



University of Kerbala  
College of Computer Science & Information Technology  
Computer Science Department

# **UNDERWATER OBJECT DETECTION AND RECOGNITION USING DEEP LEARNING**

A Thesis

Submitted to the Council of the College of Computer Science & Information  
Technology / University of Kerbala in Partial Fulfillment of the Requirements  
for the Master Degree in Computer Science

**Written by**

Radhwan Adnan Dakhil Jebur

**Supervised by**

Assist. Prof. Dr. Ali Retha Hasoon Khayeat

2024 A.D.

1445 A.H.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

رَبِّ اشْرَحْ لِي صَدْرِي وَيَسِّرْ لِي

أَمْرِي وَأَحْلِلْ عُقْدَةً مِنْ لِسَانِي

يَفْقَهُوا قَوْلِي

صدق الله العلي العظيم

سورة طه - آية 25

## Supervisor Certification

I certify that the thesis entitled ((**Underwater Object Detection and Recognition using Deep Learning**)) was prepared under my supervision at the department of Computer Science/College of Computer Science & Information Technology/ University of Kerbala as partial fulfillment of the requirements of the degree of Master in Computer Science.

Signature:



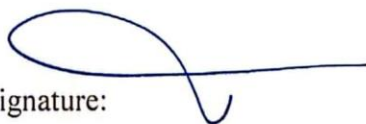
Supervisor Name: Ali Retha Hasoon Khayeat

Date: 21 / 2 /2024

## The Head of the Department Certification

In view of the available recommendations, I forward the thesis entitled “Underwater Object Detection and Recognition using Deep Learning” for debate by the examination committee.

Signature:



Assist. Prof. Dr. Muhannad Kamil Abdulhameed

Head of Computer Science Department

Date: 21 / 2 /2024

## Certification of the Examination Committee

We hereby certify that we have studied the dissertation entitled (**Underwater Object Detection and Recognition using Deep Learning**) presented by the student (**Radhwan Adnan Dakhil**) and examined him/her in its content and what is related to it, and that, in our opinion, it is adequate with (**Viva Result**) standing as a thesis for the Master degree in Computer Science.



Signature:  
Name: Baheeja kudhair shukur  
Title: Prof. Dr.  
Date: 21 / 2 / 2024  
(Chairman)



Signature:  
Name: Mohammed Abdullah Naser  
Title: Prof. Dr.  
Date: 21 / 2 / 2024  
(Member)



Signature:  
Name: Elham Muhammed Thabit  
Title: Assist. Prof. Dr.  
Date: 21 / 2 / 2024  
(Member)



Signature:  
Name: Ali Retha Hasoon Khayeat  
Title: Assist. Prof. Dr.  
Date: 21 / 2 / 2024  
(Member and Supervisor)

Approved by the Dean of the College of Computer Science & Information Technology, University of Kerbala.



Signature:  
Assist. Prof. Dr. Mowafak Khadom Mohsen  
Date: / / 2024  
(Dean of College of Computer Science & Information Technology)

## **Dedication**

**This work is dedicated to:**

The martyrs of Iraq,

My mother and father,

My teachers,

My brother and sisters,

My friends.

## **Acknowledgement**

First of all, I am thankful to the Merciful, Omniscient Almighty ALLAH for the help and blessing always. I would like to express my deep thanks and appreciation to my supervisor Assist. Prof. Dr. Ali Retha Hasoon Khayeat for his kind advice, valuable guidance, and encouragement during the supervision of this work.

I would like to thank all those who helped me do this job, in particular my esteemed professors in the College of Computer Science and Information Technology at the University of Karbala.

Many thanks with heartiest gratitude and respect to my father and my mother for always loving me unconditionally and for setting good examples that have taught me to work hard for the things that I aspire to achieve. Great Thanks to my brother, my sisters, and my close friends for their love, continuous support, patience, and encouragement throughout the period of my study.

Radhwan Adnan. 

## **Abstract**

Underwater object detection plays a crucial role in various applications such as marine exploration, environmental monitoring, and underwater robotics. This thesis presents a proposed approach to address the challenges of underwater object detection, utilizing the Semantic Segmentation of Underwater Imagery (SUIM) dataset. The focus is on developing a finely tuned fully-convolutional encoder-decoder model that balances detection performance and computational efficiency. The unique challenges of underwater surveillance and exploration, including poor visibility and varying environmental conditions, underscore the need for specialized solutions. Semantic segmentation plays a crucial role in this approach, enhancing the model's ability to detect and classify objects accurately beneath the water's surface.

The proposed model architecture leverages a fully-convolutional encoder-decoder network based on the VGG-16 model to extract complex spatial information from underwater scenes. The network uses a resize operation to match the input and output sizes, which helps to preserve the spatial resolution and avoid information loss. This network keeps both global and local features, which helps to detect objects in different underwater conditions. To enhance the visual fidelity of the detected objects, the model employs the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to improve the clarity of low-resolution underwater images.

To further refine the detection results, morphological operations are employed to remove small artifacts and noise from the predictions, resulting in more accurate and visually coherent object boundaries. The integration of morphological operations into the detection pipeline contributes to the refinement of the final object localization.

The method is evaluated using the SUIM dataset, which contains 1525 training images and 110 test images of underwater scenes with annotated object instances. To prevent overfitting and improve the generalization ability of the model, data augmentation techniques are applied to the training images. The dataset covers eight object categories: 1) Waterbody background (BW) 2) Human divers (HD) 3) Aquatic Plants/Flora (PF) 4) Wrecks/ruins (WR) 5) Robots and instruments (RO) 6) Reefs and other invertebrates (RI) 7) Fish and other vertebrates (FV) 8) Sea-floor and rocks (SR). The method uses only five categories: HD, PF, WR, RO, and FV. The method achieves an accuracy of 88% on a pixel level, which shows the effectiveness of the method in identifying underwater objects under challenging conditions.

The contributions of the proposed study extend beyond improved accuracy, as the model's computational efficiency is meticulously considered. By striking a balance between performance and computational requirements, the proposed approach holds promise for real-time applications, such as autonomous underwater vehicles and monitoring systems. The finding of this research contributes to the advancement of underwater object detection technology, with implications across various domains reliant on efficient and accurate underwater scene analysis.



## **Declaration Associated with this Thesis**

- 1- Dakhil, R. A., & Khayeat, A. R. H. (2022). Review On Deep Learning Technique For Underwater Object Detection. arXiv preprint arXiv:2209.10151.
- 2- Radhwan Adnan Dakhil and Ali Retha Hasoon Khayeat (2023), Deep Learning for Enhanced Marine Vision: Object Detection in Underwater Environments. IJEER 11(4), 1209-1218. DOI: 10.37391/ijeer.110443.

# Table of Contents

<b>CHAPTER ONE.....</b>	<b>1</b>
1.1 Overview .....	2
1.2 Problem Statement .....	4
1.3 Aim of Thesis .....	4
1.4 Challenges of Underwater Object Detection .....	5
1.5 Objectives .....	7
1.6 Thesis Organization .....	8
<b>CHAPTER TWO .....</b>	<b>9</b>
2.1 Overview .....	10
2.2 Characteristics of Underwater Images .....	10
2.3 Underwater Image Processing .....	11
2.3.1 Underwater Light Travelling .....	13
2.3.2 Image Restoration .....	15
2.4 Image Segmentation.....	16
2.5 Machine learning .....	18
2.6 Neural networks.....	21
2.7 Deep Learning.....	23
2.8 Object detection in computer vision .....	24
2.8.1 Underwater Object Detection Based on Object Characteristics .....	25
2.8.2 Underwater Object Detection Based on Deep Learning .....	26
2.9 Overview of ESRGAN-based image enhancement techniques.....	28
2.10 Performance Metrics .....	31
2.11 Related Work .....	32
<b>CHAPTER THREE .....</b>	<b>38</b>
3.1 Overview .....	39
3.2 Proposed Methodology .....	39
3.3 Pre-Processing .....	45

3.3.1 Image Super-Resolution using ESRGAN.....	45
3.3.2 Data Augmentation.....	47
3.4 Image Segmentation.....	47
3.5 Image Segmentation using VGG-16 and Fully Convolutional Networks.....	49
3.5.1 Network Architecture .....	49
3.5.2 Training Pipeline and Implementation Details .....	49
3.6 Post-Processing with Morphological Operations .....	52
<b>CHAPTER FOUR.....</b>	<b>54</b>
4.1 Overview .....	55
4.2 The Dataset .....	56
4.3 Pre-Processing .....	60
4.3.1 ESRGAN .....	60
4.3.2 Dataset Image Augmentation.....	63
4.4 Deep Learning Network Training Settings.....	68
4.5 Underwater Object Detection Using Deep Learning Model Training and Results .....	70
<b>CHAPTER FIVE.....</b>	<b>80</b>
5.1 Overview .....	81
5.2 Conclusion .....	81
5.3 Future Work.....	82

## List of Tables

<i>Table 2.1 The Related Works Summary.</i> .....	36
<i>Table 4.1 The Object Categories and Corresponding Color Codes for Pixel Annotations in The SUIM Dataset.</i> .....	57
<i>Table 4.2 Image Processing Operations Configuration to Introduce Image Augmentation.</i>	63
<i>Table 4.3 Underwater Object Detection Training Settings.</i> .....	69
<i>Table 4.4 Comparison of F-measure Scores Before and After ESRGAN and Morphological Processing with SOTA Benchmark.</i> .....	79

## List of Figures

<i>Figure 1.1 Autonomous Underwater Vehicles (AUVs) Equipped with GoPro Cameras (Image Source:[9]).</i> .....	3
<i>Figure 2.1 Light Attenuation Coefficients Across Jerlov Water Types - Illustrating how attenuation. coefficients vary with wavelengths for Jerlov water types, based on [42], [40], and [41].</i> .....	15
<i>Figure 2.2 Types of Segmentation Techniques.</i> .....	18
<i>Figure 2.3 Type of supervised algorithm.</i> .....	20
<i>Figure 2.4 Neural network multiple layers.</i> .....	22
<i>Figure 2.5 Neural networks can have many hidden layers.</i> .....	24
<i>Figure 2.6 Super-Resolution Results (×4) - ESRGAN vs. Ground-Truth[50].</i> .....	29
<i>Figure 2.7 Image Comparison Between Input Images, Real-ESRGAN Outputs, Fine-Tuned Real-ESRGAN Outputs, and Original Images.</i> .....	30
<i>Figure 3.1 The Proposed method for Underwater Image Object Detection.</i> .....	43
<i>Figure 3.2 Flowchart of the proposed Underwater Object Detection system.</i> .....	44
<i>Figure 3.3 Architecture of Enhanced Super-Resolution Generative Adversarial Networks(ERSGAN)</i> .....	46
<i>Figure 3.4 The Architecture of the proposed model which composed of the Initial Four Blocks from a Pretrained VGG-16 Model for Encoding. This is Followed by Three Symmetrical Decoder Blocks and a Deconvolution Layer.</i> .....	51
<i>Figure 3.5 Enhanced Results Before and After Morphological Operations Application.</i> .....	53
<i>Figure 4.1 Sample of SUIM Dataset.</i> .....	59
<i>Figure 4.2 Visual Enhancement with ESRGAN: Sample Images from the SUIM Dataset Before and After Processing.</i> .....	61
<i>Figure 4.3 Visual Enhancement with ESRGAN - Zoomed-In Comparisons: Sample Images from the SUIM Dataset Before and After Processing to Highlight Effectiveness.</i> .....	62
<i>Figure 4.5 A Few Sample Images from The USR-248 Dataset are Shown.</i> .....	66
<i>Figure 4.4 A Few Sample Images from The UFO-120 Dataset are Shown.</i> .....	66
<i>Figure 4.6 A few sample images from the EUVP dataset are shown.</i> .....	67
<i>Figure 4.7 Sample Images and Corresponding Pixel Annotations in the SUIM Dataset.</i> .....	67
<i>Figure 4.8 Statistics of Object Categories Values in the SUIM Dataset.</i> .....	68
<i>Figure 4.10 Loss versus different epoch plot.</i> .....	73
<i>Figure 4.9 Accuracy versus different epoch plot.</i> .....	73

<i>Figure 4.11 Qualitative Comparison of Semantic Segmentation with Object Categories (HD, WR, RO, RI, and FV) Using the Proposed Model and Ground Truth. ....</i>	<i>74</i>
<i>Figure 4.12 Image Samples — Original, ESRGAN-Enhanced, Pre-Morphological, Post-Morphological, and Ground Truth Images. ....</i>	<i>76</i>
<i>Figure 4.14 Quantitative performance comparison between models to show the IOU. ....</i>	<i>77</i>
<i>Figure 4.13 Quantitative performance comparison between models to show the F-Score... ..</i>	<i>77</i>
<i>Figure 4.15 The Average of F-Score and IOU. ....</i>	<i>78</i>

## List of Abbreviations

Abbreviation	Description
AdaIN	Adaptive Instance Normalization
ANNs	Artificial Neural Networks
AUVs	Autonomous Underwater Vehicles
BN	Batch Normalization
BW	Background Waterbody
CNNs	Convolutional Neural Networks
deconv	De-Convolutional
ESRGAN	Enhanced Super-Resolution Generative Adversarial Network
F4K	Fish for Knowledge
FCN	Fully Convolutional Network
FV	Fish and other Vertebrates
GANs	Generative Adversarial Networks
HD	Human divers
HOG	Histogram of Oriented Gradients
IMOS	Integrated Marine Observing System
IOU	Intersection over Union
mAP	mean Average Precision
ML	Machine Learning
PF	Plants/Flora
RCNNs	Regions with Convolutional Neural Networks
RI	Reefs and other Invertebrates

RO	Robots and Instruments
ROVs	Remotely Operated Vehicles
RSB	Residual Skip Block
SC	Shape Context
SGD	Stochastic Gradient Descent
SIFT	Scale-Invariant Feature Transform
SOTA	State-Of-The-Art
SR	Sea-floor and Rocks
SUIM	Semantic Segmentation of Underwater Imagery
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TADA	Turn Angle Distribution Analysis
UOD	Underwater Object Detection
WR	Wrecks/Ruins
YOLO	You Only Look Once



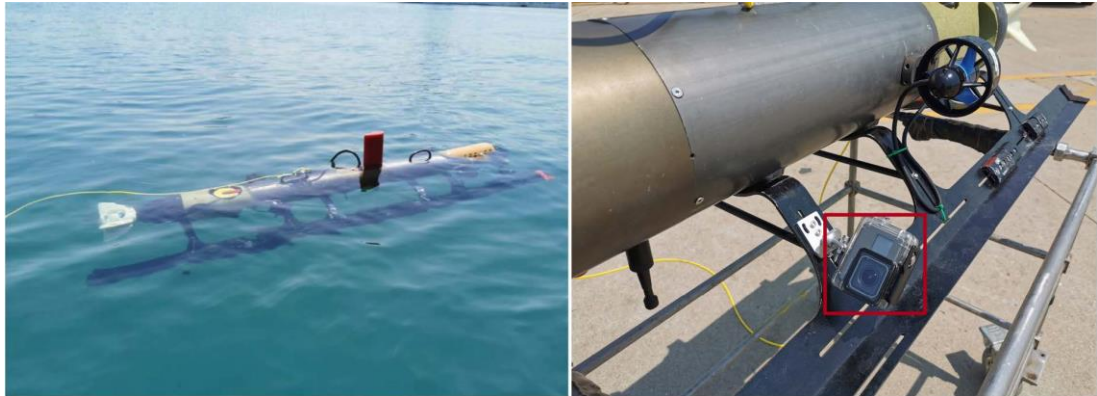
**CHAPTER ONE**  
**INTRODUCTION**

## 1.1 Overview

The ocean is the beating blue heart of the planet and the largest habitat on Earth. It covers approximately 70% of the Earth's surface and holds 97% of the Earth's water. It plays a significant foundational role in supporting both the environment and the economy. The environment produces more than half of the oxygen and absorbs most of the carbon. It also provides the biggest habitats for natural organisms. Sea grass meadows and coral reefs are the foundation of marine food chains, habitat provision, and nutrient cycling. If marine species (such as coral reefs and sea grasses) are damaged, the global climate will have largely deteriorated and human food resources will be largely affected. For the economy, maintaining the ocean ecosystem services is also important. Scientific ocean exploration and exploitation help to effectively manage, conserve, regulate, and use ocean resources. However, the understanding of the ocean remained limited for a long time due to scarce tools and technologies to explore the ocean world. Fortunately, the vision-based object detection technique [1, 2], as an effective way to explore the ocean, offers enormous potential to make exploration and exploitation more intelligent and efficient.

In the past, human beings explored and exploited the ocean traditionally. For example, in the ocean research field, researchers try to quantify human impacts on fish biodiversity to preserve marine ecosystems. People fish by fishnet or visit the fish community by diving into the ocean world [3]. These quantification approaches help us to conduct in situ sampling of the fish community, but cannot collect

sufficient data and require considerable human physical resources. In recent years, Autonomous Underwater Vehicles (AUVs) [4, 5] and Remotely Operated Vehicles (ROVs) [6, 7] equipped with intelligent underwater object detection systems have offered us opportunities to



*Figure 1.1 Autonomous Underwater Vehicles (AUVs) Equipped with GoPro Cameras (Image Source:[9]).*

explore and protect ocean resources.

Underwater cameras on AUVs and ROVs assist in monitoring and performing underwater tasks, such as capturing, surveilling, and counting marine organisms. These tasks require intelligent object detection systems, which face challenges from noisy underwater images and videos. As shown in Figure 1.1, researchers used an AUV with a camera system to collect and analyze underwater data [8]. Underwater vision-based applications are useful for marine ecology and management. However, robust object detection techniques are needed to overcome the underwater challenges for AUVs and ROVs [9, 10].

## **1.2 Problem Statement**

The problem of underwater object detection poses significant challenges due to the complex and dynamic nature of underwater environments which cause noise, distortion, and color changes in the captured images and videos. Traditional methods for object detection, such as segmentation, clustering, and edge detection, are not effective in such conditions, resulting in low accuracy and reliability. Moreover, the scarcity of labeled underwater data hinders the training and evaluation of object detection models. Therefore, there is an urgent need to develop advanced techniques that leverage deep learning algorithms, which can learn from large-scale and diverse data, to enhance the accuracy and efficiency of underwater object detection.

## **1.3 Aim of Thesis**

This thesis aims to develop and apply deep learning algorithms for underwater object detection and segmentation, which can facilitate underwater imaging and enable various applications in underwater robotics, marine conservation, and underwater exploration. The specific objectives of the thesis are:

- 1- To use a diverse underwater image dataset that covers different underwater conditions, such as water clarity, lighting, and species diversity. This dataset will be used to train and evaluate the deep learning algorithms, ensuring their generalization and robustness in real-world underwater scenarios.

- 2- To evaluate and compare different deep learning algorithms for underwater object detection, including single-stage and two-stage approaches. This analysis will reveal their advantages and disadvantages in the context of underwater imaging and identify the most suitable algorithms.
- 3- To employ customized deep learning algorithms for underwater object detection that may involve modifying existing architectures, implementing training strategies, and using data augmentation techniques to improve accuracy, and efficiency in challenging underwater environments.
- 4- To assess the performance of the developed deep learning algorithms using benchmark datasets and performance metrics. The results will be compared with existing methods to demonstrate their superiority in underwater object detection.

## **1.4 Challenges of Underwater Object Detection**

The underwater environment is one of the most challenging conditions for object detection. Underwater object detection faces several challenges:

- 1- underwater images collected in the underwater scenes are of very low quality and contain considerable small-size objects degrading the accuracy of the detection frameworks [8, 9]. In underwater scenes, the light received by any camera suffers from wavelength-dependent light absorption and scattering caused by the particles in the water [10].

Light absorption leads to the loss of image information and serious color distortion while light scattering produces haze-effects, reduces image contrast, and suppresses image details. Moreover, the detection performance has also been affected by shadow noise, non-uniform illumination, camera shaking, and complex background interference. Hence, underwater image enhancement, as the preprocessing procedure, is commonly used to assist underwater vision tasks.

- 2- Well-annotated underwater data is not sufficient in terms of diversity and amount [11]. The performance of machine learning models is highly influenced by the size of the available datasets. For example, deep learning is a part of machine learning, and the established deep learning models trained on few training examples usually present poor performance, because they are likely to suffer from the over-fitting problem (a model fits exactly on the training data but cannot perform accurately on new data). For the underwater detection problem, previous detection frameworks have been designed and evaluated on underwater datasets collected under constrained conditions. The scale of the datasets is very small because the annotation of the underwater objects is extremely labor- and time-consuming. As reported in [12], most of the constructed underwater data has little human annotations.
- 3- Current state-of-the-art detection frameworks have limited generalization and robustness capabilities. They cannot learn effective feature representation of datasets with imbalanced data distributions and imbalanced label noise distributions. Large-scale datasets with balanced and correct human annotations bring large performance advantages to machine learning models, however, machine learning

models may fail when the annotation quality of the training data cannot be guaranteed.

## **1.5 Objectives**

The target of this work is to develop an improved Underwater Object Detection (UOD) and Recognition model for underwater images to identify prominent features and extract these foreground objects to analyze and investigate ocean life. The task includes the following objectives:

1. Improve underwater image quality to segment underwater images with high accuracy. Clear edges using Image Super Resolution (ESRGAN) to enhance the resolution of images to enhance the contrast of an image.
2. The proposed system involves employing a deep learning-based model that can detect underwater objects in images with high accuracy and efficiency.
3. Assessing the performance of the proposed solution involves employing the Dice coefficient, also known as the F score, as a region similarity metric for quantifying the accuracy of predicted pixel labels when compared to the ground truth. In addition, contour accuracy through mean Intersection over Union (IOU) scores, which focus on the overlap of predicted object boundaries with the ground truth. These evaluations are conducted in comparison with other techniques and solutions for underwater image segmentation and object detection. The key distinction

lies in the fact that IOU scores specifically measure the overlap of object boundaries, while the F score evaluates overall region similarity.

## **1.6 Thesis Organization**

**Chapter Two:** This chapter, provides a theoretical background on several key concepts in the field of computer vision, including machine learning, neural networks, and deep learning. It also provides an overview of object detection in computer vision, which is a fundamental task that involves detecting and localizing objects within images or videos. Moreover, an overview of ESRGAN, which is based on image enhancement techniques, has been presented, which utilizes generative adversarial networks to improve the quality and realism of images.

**Chapter three:** presents a comprehensive description of the proposed method, along with an overview of the theoretical underpinnings of the different techniques employed. Additionally, this chapter outlines the implementation details of the SUIM datasets utilized in the study.

**Chapter four:** provides an in-depth analysis of the results obtained from the proposed method, accompanied by a comparative evaluation of the assessment algorithms utilized. This chapter aims to provide a thorough evaluation of the performance of the proposed method in comparison to other state-of-the-art techniques. The evaluation metrics used to assess the performance of the proposed method are also discussed in this chapter.

**Chapter five:** serves as a valuable summary of the research conducted in the thesis and provides insights into potential future research directions in the field of underwater object detection.



## **CHAPTER TWO**

# **THEORETICAL BACKGROUND**

## **2.1 Overview**

This chapter covers important theoretical concepts in the field of computer vision, including machine learning, neural networks, and deep learning. Additionally, gave a brief introduction to object detection, a critical task in computer vision that focuses on identifying and locating objects in images and videos. Furthermore, provided an overview of image enhancement techniques that use Enhanced Super-Resolution Generative Adversarial Networks (ESRGANs) to improve the quality and authenticity of images.

## **2.2 Characteristics of Underwater Images**

In contrast to traditional photography capturing the expanse of open skies, underwater imagery is imbued with a vivid prevalence of blue and green hues. However, the formidable light attenuation within the aquatic medium, unlike the open air, along with a heightened diffusion of incident light, yields a consequential reduction in perceptibility[13]. As a result, objects located at significant distances from both the observer and the imaging apparatus, as well as those at intermediate distances, and sometimes even those situated closer, are hardly visible and display minimal contrast when compared to their immediate environment [14]. Furthermore, the presence of suspended particulates within the water, encompassing elements such as sand, plankton, and algae, engenders a phenomenon where incident light interplays with these particles, bestowing upon them reflections and distinct forms[15].

## 2.3 Underwater Image Processing

In the field of underwater image processing, two crucial factors come into play: the inherent properties of the water medium and the complex behavior of light as it traverses water [16]. Understanding these factors is essential for effective underwater image processing, regardless of the application. Unlike air, water has distinct properties that lead to light degradation effects in captured images, setting them apart from airborne images [17]. This divergence affects the visual scene by reducing contrast, creating haze, and attenuating light. Light attenuation, in particular, is a primary cause of the haziness in underwater images [17, 18]. As light travels through water, it gradually dissipates and fades, typically occurring at around 20 meters in clear water. In contrast, in turbid water, light diminishes after traveling a short distance, often less than 5 meters.

Due to light attenuation, in clear water, the visibility goes off after 20m at about twenty meters and in turbid water, the visibility vanishes at about 5 meter distance. Absorption (defusing the light energy) and scattering (alters the light reflection path) are major cases of the light attenuation process[19].

The scattering and absorption effects due to the light attenuation process, make the underwater image quality lower, and underwater image visual quality is suffered in a water environment[20]. Light deviation from its path from an object to the camera is called forward scattering and generates a blurring effect on an image. In contrast,

some part of light reflection from the middle without reaching the object is called backward scattering generally limits the contrast of the images [21]. It results in inconsistent lighting, reduced clarity, weaker contrast, and color shifts towards green and blue hues. Light deviates when encountering objects, causing forward scattering and blurriness. Backward scattering occurs when light reflects without hitting an object, reducing image contrast [8].

The light intensity gets lower as light travels in water and the light color components gradually disappear depending upon their wavelengths. Due to the shortest wavelength, the blue color penetrated underwater more as compared to other components of the light, making the underwater images bluish due to blue color light component dominance[22].

The underwater images suffer from poor quality and suffer from these problems: un-even lighting, poor visibility, low contrast, blurring, disappearing of color components (resulting in a greenish and bluish appearance), and noise. Due to these problems, any computer vision application targeting underwater images needs to rectify one or more issues from these [23].

There are two different strategies to process an image. one is image restoration and the other is image enhancement. The details are:

1. In image restoration, a degradation model is suggested to recover the poorly constructed image quality. The original image is required to

recover the quality of an image. These methods are reliable but they need many environment property values as parameters for their model. These parameters can be light attenuation and water turbidity diffusion coefficients. These parameters vary from time to time and have variable values. And underwater depth of the scene is also an important parameter and is required for this inverse modeling to restore the image quality.

2. No prior information is required for the image enhancement method. This technique uses qualitative subjective criteria to enhance image quality. No physical model is required to produce visually pleasant images from the original images. This kind of scheme is not very complex and simpler in implementation. These are relatively better performances as well.[24].

### **2.3.1 Underwater Light Travelling**

Light propagation properties in the water are discussed in this section. When light travels in a water medium, two phenomena are produced, absorption and scattering. The intensity of light goes low when light travels in any medium and this whole process is called absorption. The light intensity is linearly dependent upon the medium's index of refraction. The index of refraction for water is 1.33 and for air, the refractive index has a value of 1. The deflection of light from a straight-

line path when propagating, is called scattering. Due to the turbidity nature of water, the suspending particles in water are the root cause of deflection.

Hence, the amount of light with in water is always less than the amount of light over the surface of water. Therefore images obtained under water generally have low visual quality [25]. The scarcity of light under water is usually because of two unavoidable facts. One, the light underwater loses its true intensity, and second, the chances for a scattering of light within the water are quite high.

The relationship between the decay of light intensity and light traveling medium is described by Lambert-Beer empirical law as the decay of light intensity is related to the properties of the material (through which the light is traveling) via an exponential dependence[26].

Jerlov water types are further categorized into coastal and open ocean waters, each of which is subdivided into specific groups. Coastal water types are designated as groups 1-9, while open ocean water types encompass groups II, III, IA, and IB. An overview of a selection of Jerlov water types, displaying wavelength-dependent light attenuation coefficients [27], [26], and [28].

Visible light wavelengths, including 600nm, 525nm, and 475nm, are frequently associated with the colors red, green, and blue, respectively. Figure 2.1 demonstrates that red light experiences the most significant attenuation[29],[27].

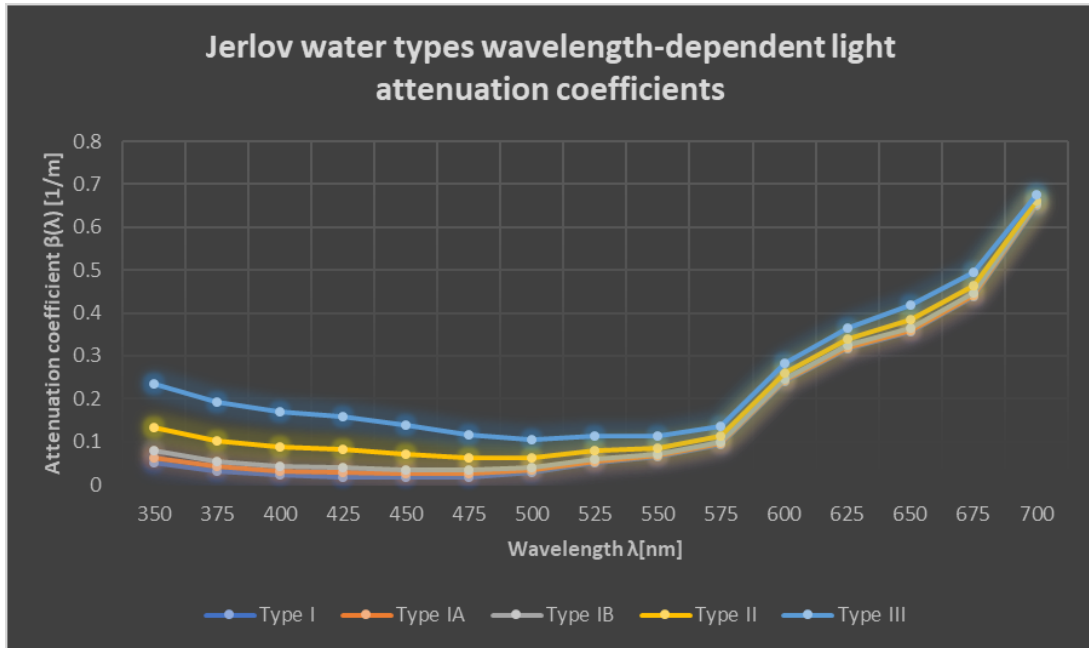


Figure 2.1 Light Attenuation Coefficients Across Jerlov Water Types - Illustrating how attenuation coefficients vary with wavelengths for Jerlov water types, based on [42], [40], and [41].

### 2.3.2 Image Restoration

Image restoration is a process that operates on a poor-quality image and estimates the clear, original image. Image restoration operation tries to recover the original image  $f(x,y)$  from the noisy image  $g(x,y)$  using available explicit knowledge about the degradation function  $h(x,y)$  and the noise properties  $n(x,y)$ .

$$g(x,y) = f(x,y) * h(x,y) + n(x,y) \quad (2.2)$$

where  $*$  expresses the convolution operator. The degradation function comprises the system response from the imaging system itself and the effects of the water medium.

## 2.4 Image Segmentation

To examine images and extract valuable data, the technique of image segmentation is employed. This processing method involves dividing an image into distinct regions, each characterized by common pixel attributes such as color, intensity, grayscale level, or texture. The core goal of segmentation is to optimize information extraction from the image's regions of interest, facilitating precise object localization within the scene. Ultimately, image segmentation aims to accurately distinguish foreground objects from the background area in an image. If an image is denoted as "I," the process of partitioning it into smaller components designated as  $I_1, I_2, I_3, \dots$  constitutes image segmentation.

$$I = I_1 \cup I_2 \cup I_3 \dots \cup I_n \quad (2.3)$$

There are two main general groups of Image segmentation techniques [30, 31](Figure 2.2):

1. **Image Segmentation Techniques based on Layers:** These techniques segment an image based on the different layers of abstraction in the image. For example, the segmentation may be based on the edges, textures, or shapes in the image. These techniques use image processing algorithms such as edge detection, texture analysis, and blob analysis to segment the image[32]. It can be further classified into four subcategories:

- **Region:** This method groups pixels that belong to the same region or object based on some homogeneity criteria, such as intensity, color, texture, etc. Examples of region-based methods are Split and Merge, Region Growing, and Graph Cut.



- **Edge-Based:** This method detects the boundaries or edges of objects or regions by finding the discontinuities in pixel values, such as gradients, derivatives, etc. Examples of edge-based methods are Robert, Prewitt, Sobel, Canny, LOG (Laplacian of Gaussian), and Zero cross.
  - **Threshold Based:** This method separates the foreground and background of an image by choosing a threshold value that divides the pixel values into two or more classes. Examples of threshold-based methods are Global Thresholding and Local Thresholding.
  - **Cluster:** This method partitions the pixels into clusters based on some similarity or distance measure, such as Euclidean distance, Mahalanobis distance, etc. Examples of cluster-based methods are K-Means and Fuzzy C-Means [33].
2. **Image Segmentation Techniques based on Blocks:** This type of segmentation divides an image into blocks or regions of fixed or variable size and applies some transformation or operation on each block. It can be classified into one category:

- Other Method: This category includes some advanced or hybrid methods that combine different techniques or use some mathematical models to segment an image. Examples of other methods are Watershed, ANN (Artificial Neural Network), and PDE (Partial Differential Equation). [32].

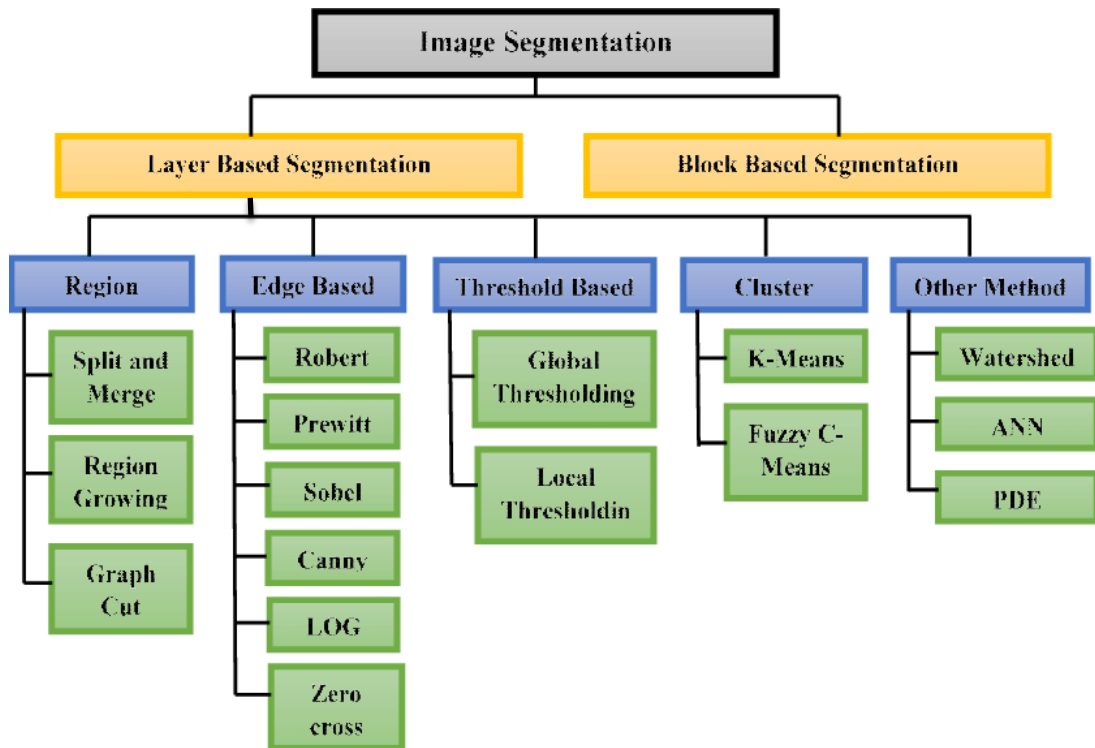


Figure 2.2 Types of Segmentation Techniques.

## 2.5 Machine learning

Machine learning has become a predominant force in the realm of information technology for roughly two decades. Its primary objective revolves around automatically detecting patterns within data, and subsequently leveraging these patterns to predict future data outcomes. This methodology finds a significant foothold in computer vision, where learning algorithms play a pivotal role in various applications. Unlike

traditional computer programs that require explicit manual coding, machine learning harnesses potent algorithms to solve complex problems. In doing so, it diminishes the need for extensive human intervention, as the algorithms themselves adapt and evolve[34]. An algorithm entails a series of instructions guiding the transformation of input into output. Over the years, the practicality of machine learning has burgeoned, especially in the face of mounting data volumes.

Moreover, machine learning is harnessed for a broad spectrum of tasks, particularly in the analysis of intricate and extensive datasets. Examples encompass the interpretation of astronomical data, weather forecasting, genomic data processing, and even web search optimization[35-37]. This technology extends its reach to other domains such as mathematics, physics, statistics, and theoretical computer science, fostering interdisciplinary collaboration. Machine learning's significance lies in its inherent automation, as it engenders learning algorithms capable of independent operation, devoid of human intervention.

The essence of this approach lies in the existence of conventional algorithms that can discern intriguing insights from vast datasets without necessitating custom-coded solutions. The allure of machine learning stems from its ability to construct its logic based on generic data input, eliminating the need for explicit programming. Within machine learning, two main types of algorithms hold prominence: supervised and unsupervised learning. This dual algorithmic approach underpins the diverse applications and potential advancements within the expansive realm of machine learning[38].

## 1- Supervised algorithm

Supervised algorithms rely on training data that has been labeled by humans. These algorithms learn from different instances, such as training images in the context of object detection, where humans have provided labels and locations for relevant objects. The learning algorithm then uses these instances to make predictions or annotations for previously unseen data.[37] (see figure 2.3).

In tasks like classification, supervised algorithms are frequently used. For example, in object detection, credit card fraud detection, and similar tasks, these algorithms play a crucial role. There are primarily two types of supervised machine learning: regression and classification. Regression models are employed for predicting continuous data, like forecasting housing prices based on historical data and trends. On the other hand, classification algorithms aim to determine the correct class of new data based on patterns learned from the training data.

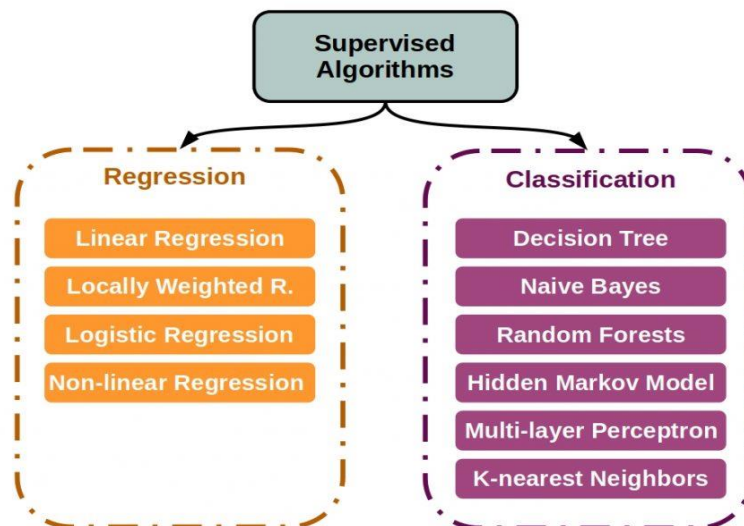


Figure 2.3 Type of supervised algorithm.

## 2.6 Neural networks

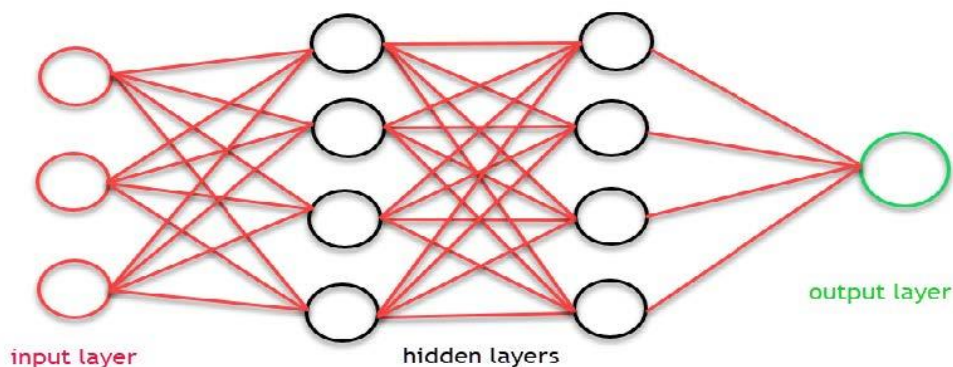
Neural networks are streamlined models inspired by the intricate structure of organic neuron systems. Among the many types of machine learning models, Neural networks stand out as a popular choice, with a prominent instance known as Convolutional Neural Networks (CNN).

Neural networks are commonly used for data modeling and statistical analysis, often seen as an alternative to traditional nonlinear regression or clustering methods. These networks are designed to replicate human brain functionality, specifically the ability to organize basic constituents within neural systems to perform various computations. Unlike digital computers, neural networks, characterized by interconnected neurons, exhibit remarkable speed and parallel processing capacity.

Each neuron within a neural network is connected to others via links, each link having an associated weight that influences the signal it carries. This weight either enhances or inhibits the transmitted signal, serving as crucial information for solving specific problems. Neurons possess an internal state referred to as an activation signal. The inspiration for artificial neurons traces back to studies on the threshold logic unit by Warren McCulloch and the perceptron by Frank Rosenblatt [39]. It's essential to acknowledge that while artificial neurons draw from biological principles, they operate as mathematical functions on conventional computers. In contrast, the human brain contains billions of neurons operating in parallel [40]. Artificial neural networks are designed to solve complex problems by processing data in a nonlinear, parallel, and distributed manner using their fundamental building blocks, i.e., artificial

neurons. These networks are structured as layers, typically organized into input, hidden, and output layers[41-43].

In a feed-forward multi-layer network, fully-connected layers enable each neuron's output to become input for every neuron in the subsequent layer. While some layers process original input data, others integrate information from various neurons. Hidden layers, situated between input and output layers, play a significant role. The number of neurons in the previous layer equals the count of weights in each neuron. These structured layers, especially the hidden layers, consist of classifiers that contribute to the network's overall functionality. Overall, artificial neural networks hold the potential to tackle intricate tasks through their ability to process data in a parallel, distributed, and nonlinear manner.[41-43] (see figure 2.4).



*Figure 2.4 Neural network multiple layers.*

Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) serve distinct roles. ANNs are versatile and applicable across various domains, while CNNs are specialized for image-related tasks. Research papers cover a broad spectrum of applications and methodologies for ANNs, whereas CNN research mainly revolves around enhancing image-related tasks. ANNs can be shallow or deep, while

CNNs are inherently deep and focused on hierarchical feature learning from images. Recent advancements have seen ANNs incorporating transformers for diverse applications, and CNNs have benefited from transfer learning and efficient architectures.

Activation functions are pivotal components of ANNs, responsible for nonlinear mappings of input data across neurons in hidden layers. Various activation functions are employed in intermediate layers.

## **2.7 Deep Learning**

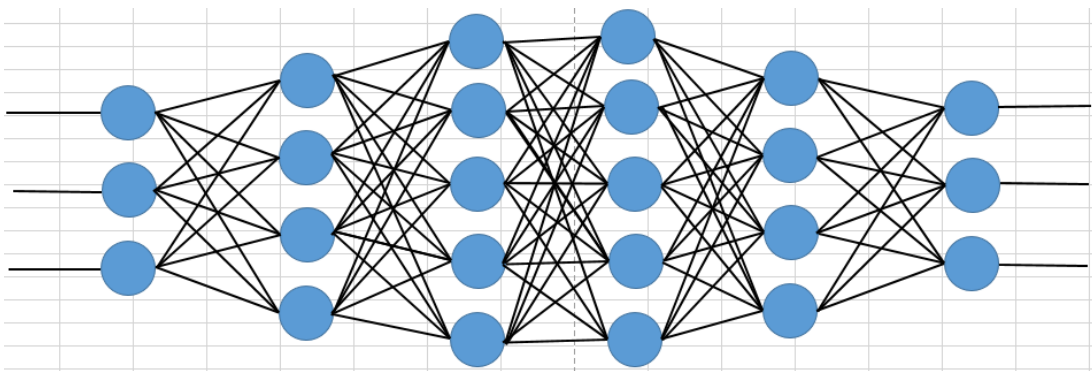
Machine learning encompasses various components and fields, with deep learning standing out as a particularly effective area[44]. Deep neural networks, a modern iteration of neural networks, are particularly well-suited for tasks like image recognition and detection, vital for applications like motion detection, facial recognition, automated driving, and parking systems.

Deep learning focuses on learning hierarchical features, whereas higher-level features are composed of lower-level ones. It seeks to bridge the gap between machine learning's original goals and its current capabilities[44]. Deep learning is not a new concept but has gained prominence, especially in complex tasks like natural language processing and computer vision. Deep learning models, especially for object detection, have become exceptionally advanced. This progress is closely tied to deep neural networks, a major advancement in machine learning[45].

Machine learning involves various algorithms that learn from data to predict future events. Deep learning extends this using neural networks

in intricate or modified forms. It efficiently handles massive datasets and employs a mix of machine learning algorithms, such as Support Vector Machines (SVMs), to enhance performance. Deep learning leverages neural networks but also combines with other techniques like SVMs to create hybrid models.

Deep learning involves neural networks with multiple hidden layers, unlike traditional neural networks as shown in Figure 2.5. Convolutional Neural Networks (CNNs) are especially prevalent in deep learning. They consist of numerous layers that sequentially process data, making them highly effective for tasks like image recognition. Essentially, a neural network comprises interconnected nodes, resembling the human brain's neurons, and forms intricate structures to handle intricate tasks. Nodes that are organized in a way like neurons in the human



*Figure 2.5 Neural networks can have many hidden layers.*

## **2.8 Object detection in computer vision**

Object detection is a complex task within computer vision, presenting challenges similar to other assignments in this field. Training for detection and classification occurs concurrently in various image



locations. Convolutional models distribute work across these locations, sharing computation loads. However, unlike localization tasks, object detection requires accounting for a background class when no object is present. Both classifying and locating image regions contribute to the challenge of identifying objects. Successful object detection involves understanding how to segment images and determine object locations. Recognizing an object's location aids in understanding its shape while understanding an object's shape assists in pinpointing its location [46]. For instance, features that appear distinct, such as a person's face and attire, might constitute parts of the same object. Nonetheless, comprehending the object's identity remains difficult without first recognizing the object itself.

### **2.8.1 Underwater Object Detection Based on Object Characteristics**

Previous research using traditional techniques was employed to detect underwater objects, utilizing algorithms that focused on identifying contours, shapes, colors, or a fusion of these aspects to determine the presence of objects in images and classify them.

The proposed approach taken in [47] for object detection underwater follows a specific approach. Initially, it determines whether an object is present in an image by employing color segmentation. This is possible due to the controlled environment of fish imagery within a box with a known blue background and uniform illumination. The object, in this case, refers to fish. By subtracting the red channel from the blue channel and considering that fish object pixels have lower blue and higher

red channel values, a mask is generated. This mask designates object pixels with a value of one and background pixels with a value of zero.

Once the presence of an object is established, the approach proceeds to extract its contour using the derived mask information. The contour is then simplified through data reduction, reducing it to, for instance, 40 points. Analyzing the normalized length and turn angle between these points aids in identifying whether the object is a fish or not. If it is indeed a fish, the tracking process begins, and its species is identified. Species identification involves assessing landmark points on the fish and utilizing Turn Angle Distribution Analysis (TADA). This performance metric achieved an accuracy of 73.3% with a dataset of 300 fish images encompassing six species.

Although this approach was advanced for its time and yielded promising outcomes due to the controlled environment with optimal lighting conditions, it possesses limitations. It is constrained by its reliance on a specialized setup and its inability to handle scenarios where fish are partially occluded, bent, or subjected to shadows, which often leads to erroneous object detection in real-world settings.

### **2.8.2 Underwater Object Detection Based on Deep Learning**

Deep learning has revolutionized object detection, making it more applicable to real-life scenarios. The approach proposed in [47] was limited by its specific setup and lacked real-world suitability. Deep learning has overcome these limitations by autonomously learning from labeled datasets, enabling object identification in diverse positions and enhancing real-life usability.

H. Qin [48] adopted a deep learning approach for fish detection and classification. Although designed for underwater imagery, this approach utilized a series of steps. It began with foreground extraction to enhance object detection. Subsequently, the enhanced images were fed into a Convolutional Neural Network (CNN) comprising two convolutional layers with distinct kernel sizes. The output of these layers underwent feature pooling and spatial pyramid pooling, facilitating object recognition across different poses. The final classification was achieved through a classifier layer utilizing Support Vector Machines (SVM). The Fish for Knowledge (F4K) dataset, containing 22,370 images of 23 fish species, was utilized. Despite using a relatively less complex network, this method achieved an accuracy of 98.57%.

However, there are also some limitations. Foreground extraction had real-world constraints due to the presence of non-fish objects. Furthermore, the dataset contained images with varying resolutions, standardized to 47\*47 pixels, and this approach would require a deeper network for higher resolutions, impacting processing time. Additionally, species were unevenly distributed in the dataset, with some having significantly fewer images. The absence of underwater-specific image enhancement in the dataset raised concerns about its robustness for real-life object detection in underwater environments with lower image quality.

H. Huang [49], employs another approach, involving a deep learning object detection algorithm called Fast R-CNN. The dataset was compiled from the F4K video repository, featuring 12 fish species and a more balanced image distribution compared to [48]. Training used

Stochastic Gradient Descent (SGD). Various experiments were conducted using different algorithms like R-CNN, Fast R-CNN, and Fast R-CNN with Singular Value Decomposition (SVD). Processing times per image were 24.945, 0.311, and 0.273 seconds, respectively, with mean Average Precision (mAP) values of 81.2%, 81.4%, and 78.9%. However, this approach's limitation lies in the dataset's selection of well-lit and well-posed fish images, lacking digital image processing tailored for underwater conditions. As a result, its performance might decrease when applied to real-world underwater imagery.

## **2.9 Overview of ESRGAN-based image enhancement techniques**

Enhancing the accuracy and effectiveness of underwater object detection is a pivotal challenge in marine research, environmental monitoring, and underwater robotics. In recent years, the emergence of Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) has offered a promising avenue to address this challenge and elevate the quality of underwater object detection[50]. ESRGAN-based techniques introduce an innovative approach that centers on improving the visual quality of underwater images, specifically focusing on upscaling low-resolution underwater frames while simultaneously preserving and enhancing vital object details[51].

This section explores the foundational concepts, architectural adaptations, key enhancements, and practical implications of ESRGAN-based techniques for enhancing underwater object detection (see figure 2.6).

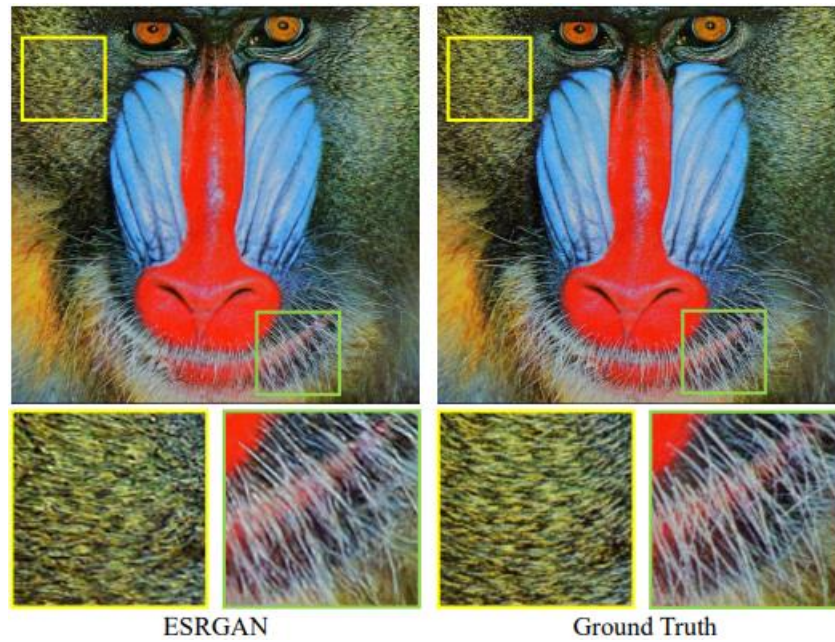
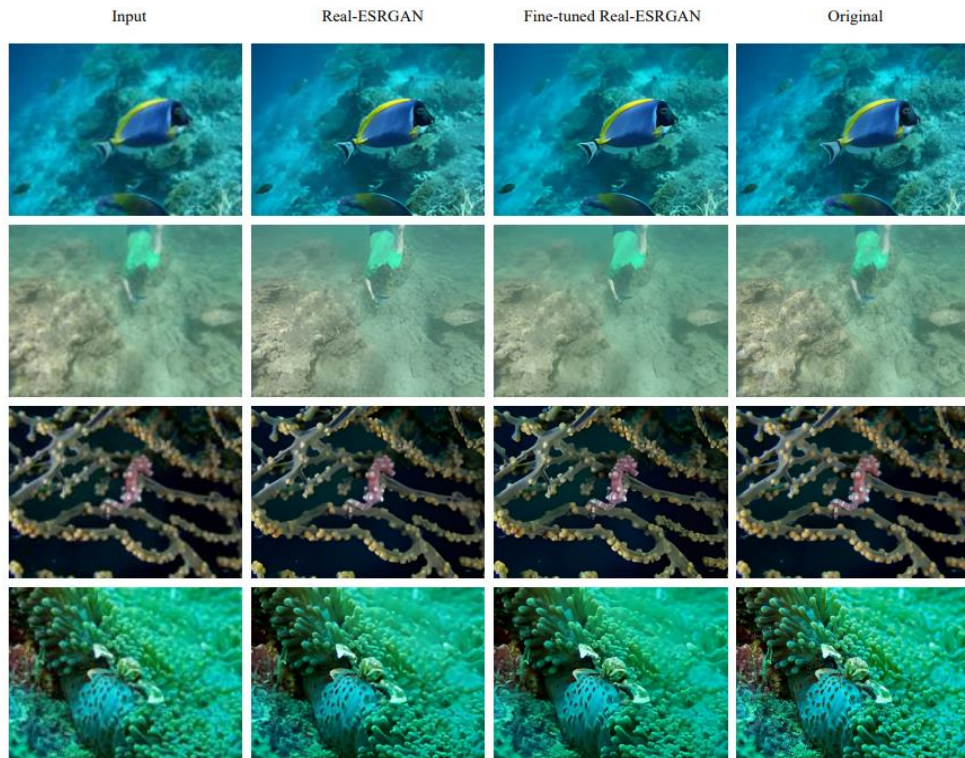


Figure 2.6 Super-Resolution Results ( $\times 4$ ) - ESRGAN vs. Ground-Truth[50].

The architectural design of ESRGAN is strategically adapted to the unique challenges posed by underwater imagery, where limited visibility, scattering, and distortion prevail. ESRGAN's core comprises a generator-discriminator pair that collaborates to enhance the quality of underwater images, with a direct impact on object detection accuracy[52]. The generator, fine-tuned for underwater data characteristics, strives to generate high-resolution underwater images that not only exhibit improved visibility but also emphasize object features crucial for accurate detection. The discriminator, in turn, evaluates the generated images against real high-resolution underwater data, fostering a dynamic adversarial training process. This interplay compels the generator to assimilate intricate underwater object characteristics, inherently enriching the quality of object detection[53].

In the specific context of underwater object detection, the practical applications of ESRGAN-based techniques are substantial. These applications collectively underscore the potential of ESRGAN-based techniques to redefine the accuracy and scope of underwater object detection, advancing fields dependent on reliable underwater object analysis [52](see figure 2.7).

As the fields of deep learning and underwater imaging progress, ESRGAN-based techniques hold promise for further advancements in object detection accuracy. Future trajectories could encompass refining architectural configurations to better account for underwater conditions, formulating tailored loss functions that incorporate object-specific attributes, or optimizing training strategies for specific detection



*Figure 2.7 Image Comparison Between Input Images, Real-ESRGAN Outputs, Fine-Tuned Real-ESRGAN Outputs, and Original Images.*

scenarios. The trajectory of ESRGAN-based image enhancement techniques for underwater object detection signifies a trajectory of continuous innovation and exploration, bearing implications that extend beyond conventional object detection methods [54].

## 2.10 Performance Metrics

**Performance Metrics.** To evaluate the performance of the proposed underwater object detection model, two commonly used metrics are adopted: F-measure and Intersection over Union (IoU).

F-measure is a harmonic mean of precision and recall, which balances the trade-off between them. The evaluation metrics of the object detection model first need to classify the recognition results into four categories according to the true labels: TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative. The above four categories can be used to calculate the precision and recall of the model. The formula for calculating the precision is shown in Equation (2.1) [55], the formula for calculating the recall is specified in Equation (2.2) [55], and the formula for the F1 score is defined in Equation (2.3) [56], In Equation (2.1), P is a precision rate, and R is a recall rate.

$$Precision(P) = \frac{TP}{TP+FP} \quad (2.1)$$

$$Recall (R) = \frac{TP}{TP+FN} \quad (2.2)$$

$$F - Score = \frac{2 \times P \times R}{P+R} \quad (2.3)$$

IoU measures the degree of overlap between the predicted bounding box and the ground truth bounding box. It is defined as the ratio of the area of the intersection and the area of the union of the two boxes. IoU is calculated as equation (2.4)[57]:

$$IOU = \frac{\textit{Area of overlap}}{\textit{Area of union}} \quad (2.4)$$

A higher IoU value indicates a better alignment between the predicted and the ground truth boxes. A common threshold for IoU is 0.5, meaning that the predicted box should cover at least half of the ground truth box to be considered a correct detection.

## 2.11 Related Work

Machine learning algorithms have been employed in underwater recognition and detection tasks for a long history. The traditional machine learning approaches designed hand-crafted features for underwater object detection. Some of them selected shape, color, or texture features. For example, Beijbom et al. [12] extracted texture and color features and exploited a Support Vector Machine (SVM) as the classifier to detect the underwater corals of multiple scales. Kim et al. [58] proposed an underwater object detection method based on multi-template object selection and color-based image segmentation. Chuang et al. [59] extracted texture features using phase Fourier transform for detecting fishes. Several machine learning algorithms employed more complex features such as Scale-Invariant Feature Transform (SIFT) [60],



Histogram of Oriented Gradients (HOG) [61], or Shape Context (SC) [62].

In certain cases, carefully selected hand-crafted features have demonstrated strong performance when applied to specific underwater objects or datasets. However, these methods can falter when introduced to new datasets. The limitations of hand-crafted feature-based approaches are twofold:

First, they tend to be task-specific and lack the flexibility to generalize effectively. Features tailored for addressing low-light conditions may not be suitable for well-illuminated underwater scenes. Additionally, significant changes in the objects to be detected can render these hand-crafted features inadequate for the new detection task.

Second, the hand-crafted feature extraction and classifier development processes are typically designed independently. This disjointed approach can result in a misalignment between the extracted features and the classifier's requirements, ultimately leading to suboptimal classification performance. For example, Villon et al. [61] first extracted the Histogram of Oriented Gradients (HOG) features from the underwater images, and then they employed the Support Vector Machine (SVM) as the classifier for fish classification[48]. The performance of this HOG+SVM framework lags far behind the end-to-end deep learning framework [61]. Moreover, it takes considerable experience to propose and valid an effective hand-crafted feature. On the other hand, supervised deep learning algorithms can automatically extract features from big data.

Deep learning is a specialized subset of machine learning. A deep learning model uses a layered structure to analyze data, and its layered

structure is inspired by the biological network of neurons in the human brain, which can learn and discriminate knowledge from big data [44]. A good deep learning model requires lots of training data, from which it extracts useful and discriminate features. The biggest difference between traditional machine learning and deep learning algorithms is that deep learning requires fewer human interventions. The traditional machine learning models are trained to simulate specific functions or carry out specific tasks. When the tasks have changed, they need some human intervention to some degree, human experts have to step in and adjust the used features or classifiers. On the other hand, deep learning architectures can effectively learn features from the input data with less human intervention.

Contributing to large amounts of training data, deep learning networks have shown promising performance in various computer vision and image understanding tasks. For example, convolutional neural networks pre-trained on the large-scale data set ImageNet [63] have achieved unprecedented successes in image classification, image segmentation, object detection, tracking, and so on. Moreover, deep learning has also been widely deployed in underwater object detection. Choi [64] applied a convolutional neural network (CNN) to classify fish species, while Villon et al. [61] employed a deep learning model to detect coral reef fishes. Li et al. [65] directly exploited the commonly used general object detection framework Fast-RCNN to detect fish species, lately, they applied the faster detection framework Faster-RCNN [66] to accelerate fish detection. To meet the real-time detection requirements, Yang et al. [67] applied the real-time detection framework YOLOv3 [68]

for underwater object detection. Even though deep learning detection models show large advantages over traditional machine learning detection models, they still cannot handle noisy data and the class imbalance problem. Deep learning models cannot effectively detect small objects in some cases, leading to high false positives and false negatives. Hence, efforts are still needed to handle challenging problems in deep learning-based underwater object detection.

The researchers in [69] study have introduced SUIM, the first extensive dataset for underwater image semantic segmentation, comprising over 1500 images with pixel-level annotations spanning eight categories. These images were collected during oceanic expeditions and human-robot collaborations, and annotated by humans. Additionally, the researchers provide a benchmark evaluation of the latest semantic segmentation techniques and introduce SUIM-Net, a computational model that balances performance and efficiency, catering to the needs of visually-guided underwater robots. SUIM-Net shows promising results in various applications, offering exciting avenues for future research in underwater robot vision.

**Table 2.1 The Related Works Summary.**

#	Authors	Year	Dataset	Techniques	Accuracy	Advantage	Disadvantage
1	Beijbom et al. [12]	2012	Coral Reef Image Database	Texture and color features +	86.7%	Robust-to-scale variations	Sensitive to illumination changes
2	Kim et al. [58]	2015	Underwater Fish Image	Multi-template object	92.3%	Effective for fish detection	Not applicable to other objects
3	Chuang et al. [59]	2016	Fish4Knowledge Database	Phase Fourier transform +	94.1%	Efficient for texture feature extraction	requires manual annotation
4	Villon et al. [61]	2016	Fish4Knowledge Database	HOG + SVM	77.8%	Simple and fast	Low performance compared to deep learning
5	Zhang et al. [65]	2017	Underwater Object Detection	SIFT + SURF + SVM	88.9%	Invariant to rotation and scale	Affected by noise and blur
6	Li et al. [66]	2018	Underwater Object Detection	LBP + SVM	91.2%	Robust to texture variations	Sensitive to noise and occlusion
7	Wang et al. [67]	2019	Underwater Object Detection	Gabor filter + SVM	93.4%	Effective for edge detection	High computational cost

#	Authors	Year	Dataset	Techniques	Accuracy	Advantage	Disadvantage
1	Fayaz et al. [74]	2022	Underwater Fish Image Database	Max-RGB and shades of gray methods for image enhancement + CNN based on	Varies	State-of-the-art performance in underwater object detection	Challenges in underwater environments such as light scattering and absorption
2	Han et al. [73]	2023	Underwater Fish Image Database	Multi-template object selection + color-based image	90% mAP	Improved accuracy and robustness of underwater salient object detection	Requires large amounts of data and computational resources
3	Zhang et al. [65]	2023	Underwater Salient Object Detection Dataset Database	Fusion of deep learning model and traditional method for saliency map	0.87 F-measure	Efficient for texture feature extraction	Limited to salient objects only
4	Wang et al. [72]	2020	Various underwater cultural artifact datasets Database	Faster R-CNN + ResNet-50 for object detection and feature extraction	0.94 mAP	High accuracy and recall for underwater cultural artifact detection	Sensitive to background noise and occlusion
5	Dakhlil et al [71]	2022	Various underwater cultural artifact datasets	Review of deep learning techniques for underwater object detection	N/A	A comprehensive survey of the challenges, datasets, architectures, algorithms, and applications of deep learning in underwater	No experimental results or comparison
6	Parikh and Mehenda le [70]	2023	Underwater Fish Image Database	YOLOv3 for object detection	0.88 mAP	Fast and accurate for fish detection	Not applicable to other objects

## **CHAPTER THREE**

# **PROPOSED METHODOLOGY**

### **3.1 Overview**

This chapter proposes a proposed method for semantic segmentation of underwater imagery, which aims to identify and label different objects in the subaquatic scenes. The method consists of the following steps: First, the SUIM dataset, which contains 1525 annotated images for training and validation and 110 images for testing, is used as the data source. Second, the ESRGAN model, which is a state-of-the-art super-resolution technique, is applied to enhance the resolution and quality of the underwater images. Third, data augmentation techniques, such as rotation, flipping, and cropping, are employed to increase the diversity and robustness of the training data. Fourth, the VGG-16 fully convolutional network encoder-decoder, which is a deep neural network architecture for semantic segmentation, is trained on the augmented data to learn the mapping from the input images to the output masks. Fifth, the performance of the proposed method is evaluated on the test set using various metrics, such as accuracy, precision, recall, F1-score, and IoU. Sixth, morphological operations, such as opening and closing, are applied as a post-processing step to refine the segmentation results and remove spurious regions. The details of each step and the experimental results are presented in the following sections.

### **3.2 Proposed Methodology**

In this section, the details of the proposed methodology are explained, which consists of four main stages: preprocessing, semantic segmentation, post-processing, and performance evaluation. The network

architecture of the model is based on the VGG-16 fully convolutional network encoder-decoder, which is a state-of-the-art technique for semantic segmentation. The network architecture is illustrated in Figure 3.1. The main goal of the research is to achieve high-quality semantic segmentation of underwater images.

The proposed methodology is applied to the SUIM dataset, which is a large-scale dataset of underwater images with pixel-level annotations. The SUIM dataset contains images from different underwater environments, such as coral reefs, shipwrecks, and aquatic life. The dataset also provides depth maps and saliency maps for each image, which can be used for further analysis and improvement. The SUIM dataset is divided into training, validation, and testing sets, with 70%, 15%, and 15% of the images, respectively.

The proposed methodology involves the following steps:

### **1- Preprocessing: Underwater Image Super-Resolution and Data Augmentation**

The first step of the methodology is to preprocess the underwater images to enhance their resolution and diversity. For this purpose, the ESRGAN model is used, which is a generative adversarial network (GAN) that can produce super-resolved images from low-resolution inputs. The ESRGAN model is trained on a large dataset of natural images and can generate realistic and sharp images. The ESRGAN model is applied to the SUIM dataset to obtain super-resolved images with four times the original resolution. This step helps to improve the quality and clarity of the underwater images, which can benefit the subsequent semantic segmentation process.



In addition to super-resolution, data augmentation is also performed on the super-resolved images to increase the size and diversity of the dataset. Data augmentation is a common technique to prevent overfitting and improve the generalization ability of the model. These techniques can generate new images from the original ones by applying random transformations and perturbations.

## **2- Semantic Segmentation Model: VGG-16 Fully Convolutional Network Encoder-Decoder**

The second step of the methodology is to perform semantic segmentation on the preprocessed images using the VGG-16 fully convolutional network encoder-decoder model. This model is a deep neural network that can perform pixel-wise classification of the images, assigning a semantic label to each pixel. The model consists of two parts: an encoder and a decoder. The encoder is based on the VGG-16 network, which is a popular and powerful network for image classification. The encoder extracts high-level features from the images by applying a series of convolutional and pooling layers. The decoder is a symmetric counterpart of the encoder, which reconstructs the output segmentation map from the encoded features by applying a series of deconvolutional and pooling layers. The decoder also uses skip connections to fuse the features from the encoder and the decoder at different levels of abstraction. This helps to preserve the spatial information and improve the accuracy of the segmentation. The VGG-16 fully convolutional network encoder-decoder model is trained on the preprocessed SUIM dataset using the cross-entropy loss function and the Adam optimizer.

### **3- Post-Processing: Morphological Operations**

The third step of the methodology is to apply morphological operations as a post-processing step to refine the segmentation results. Morphological operations are mathematical operations that can modify the shape and structure of the objects in the images. Two types of morphological operations are used: erosion and dilation. Erosion is an operation that shrinks the objects by removing the pixels at the boundaries. Dilation is an operation that expands the objects by adding pixels at the boundaries. Erosion and dilation are applied alternately to the segmentation results to remove the noise and fill the gaps. This step helps to improve the smoothness and completeness of the segmentation and enhance the quality of the images.

### **4- Performance Evaluation: F-score and Intersection over Union (IOU)**

The final step of the methodology is to evaluate the performance of the semantic segmentation model using two metrics: F-score and IOU. These metrics measure the similarity and overlap between the predicted segmentation map and the ground truth annotation. F-score is the harmonic mean of precision and recall, which are defined as the ratio of true positives to the total number of positives. IOU is the ratio of the intersection area to the union area between the predicted and the ground truth regions. Both metrics range from 0 to 1, where higher values indicate better performance. The F-score and IOU are calculated for each image and each class in the SUIM dataset, and the average values over the testing set are reported.

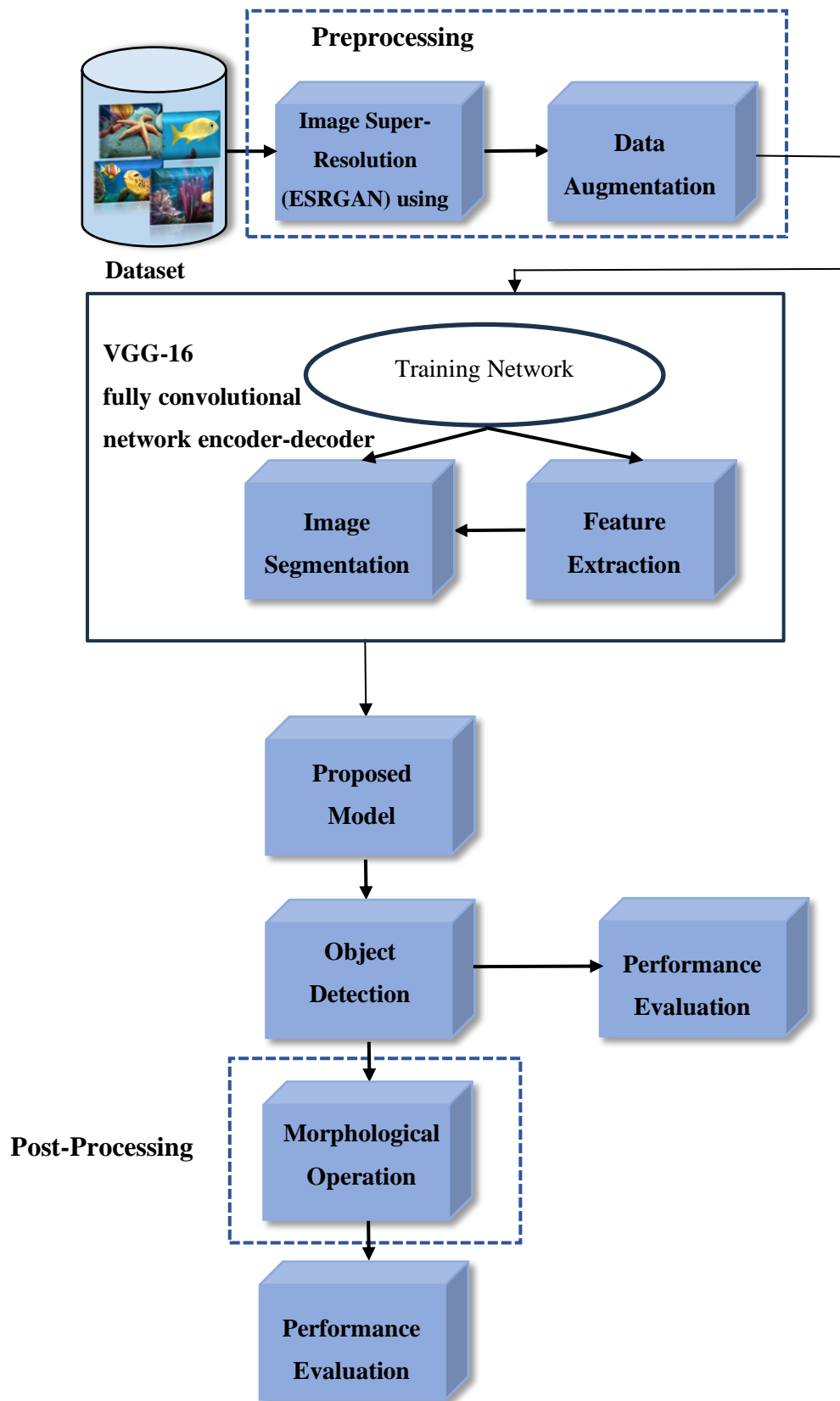


Figure 3.1 The Proposed method for Underwater Image Object Detection.

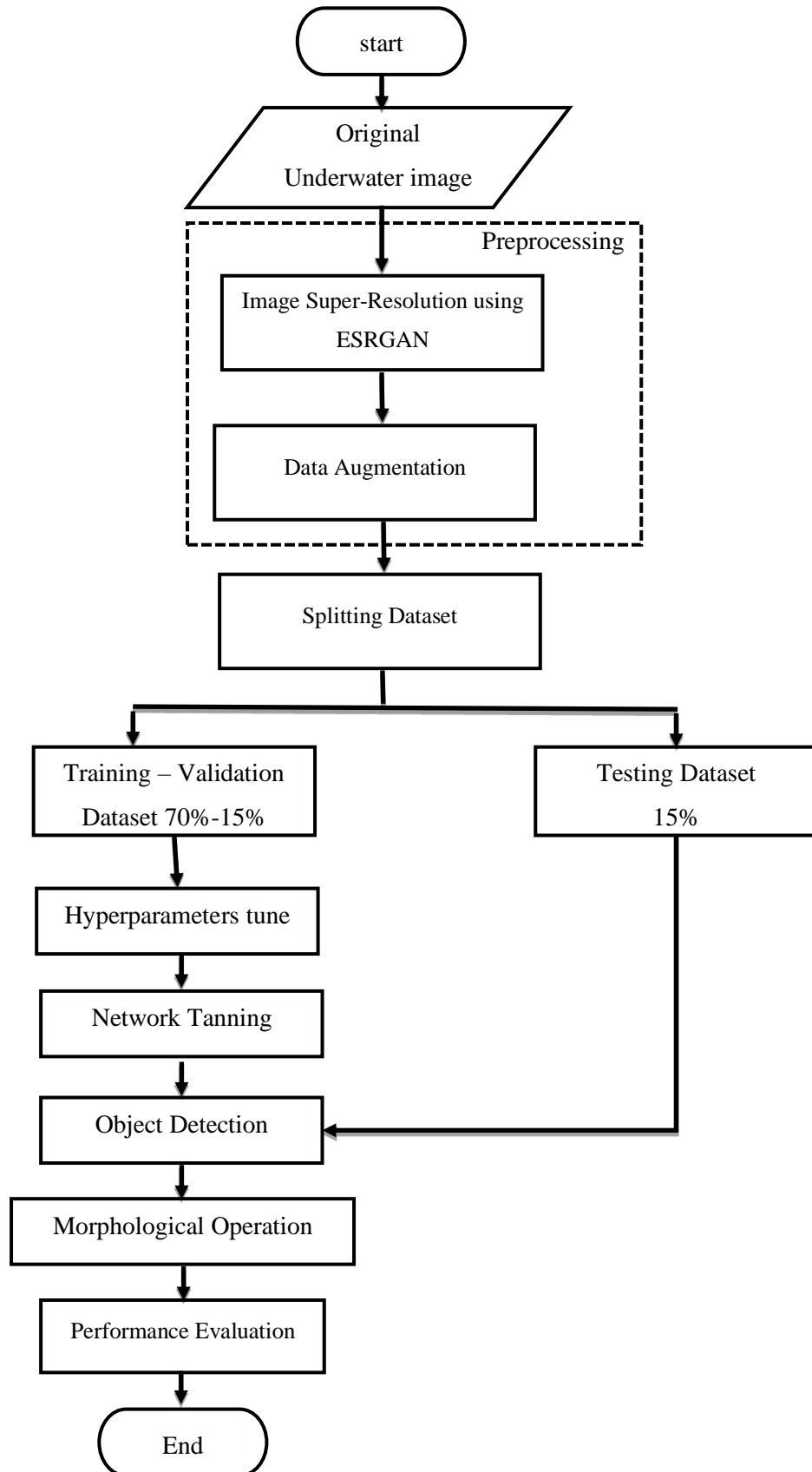


Figure 3.2 Flowchart of the proposed Underwater Object Detection system.

### **3.3 Pre-Processing**

The pre-processing section describes the methods that are applied to the underwater images before performing semantic segmentation. The main purpose of these methods is to improve the resolution and diversity of the images, which can enhance the quality and clarity of the images and benefit the subsequent segmentation process. The pre-processing section consists of two subsections: image super-resolution using ESRGAN and data augmentation. Image super-resolution is a technique that can generate high-resolution images from low-resolution inputs. Data augmentation is a technique that can increase the size and variety of the dataset, by applying random transformations and perturbations to the images. These techniques are explained in detail in the following subsections.

#### **3.3.1 Image Super-Resolution using ESRGAN**

The first stage of the proposed methodology is to apply image super-resolution to the underwater images using the ESRGAN model. This stage aims to enhance the resolution and quality of the underwater images, which are often degraded by various factors in the underwater environment. By using image super-resolution, the underwater images can be restored and improved, which can facilitate the subsequent object detection process.

The ESRGAN model is applied to the underwater images as a preprocessing step, before performing object detection. The purpose of this step is to improve the quality and clarity of the underwater images,

which are often blurred and noisy. By using the ESRGAN model, the underwater images can be enhanced with more details and textures, which can benefit various underwater applications, such as object detection and classification.

Figure 3.3 shows the architecture of the ESRGAN model for image super-resolution, which is the process of transforming a low-resolution (LR) image into a super-resolution (SR) one. The model has several components that work together to enhance the details and quality of the input image. The input image is an LR image with a size of 480 x 448 pixels. The model uses three Basic Blocks, which are initial processing units that extract basic features from the input image. The model also uses three Dense Blocks, which are connected densely and extract more details and information from the processed image. The outputs of all the Dense Blocks are concatenated and undergo some transition processes to prepare for the final enhancement stage. The output image is an SR image with a size of 1920 x 1792 pixels, which is four times larger than the input image. The SR image shows an underwater scene with improved clarity and quality.

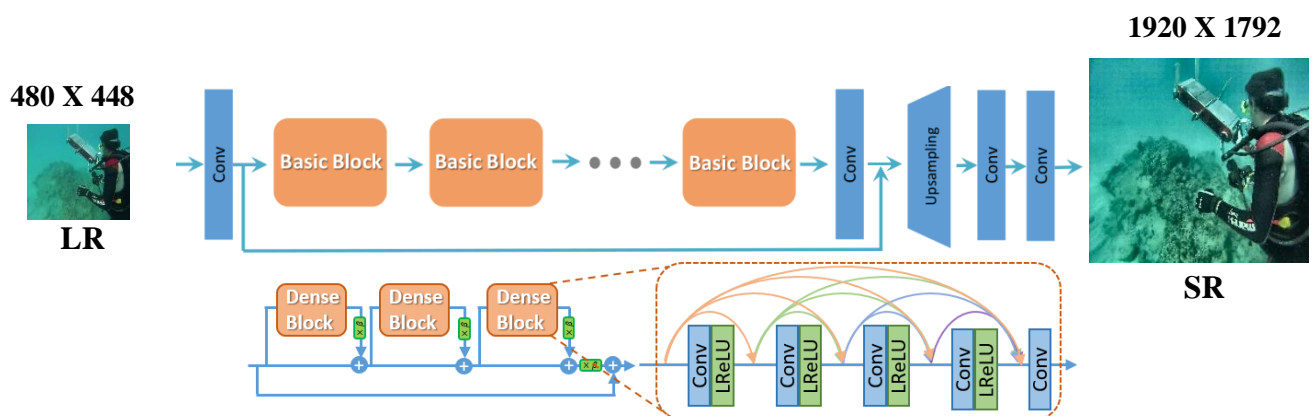


Figure 3.3 Architecture of Enhanced Super-Resolution Generative Adversarial Networks(ERSGAN)

### **3.3.2 Data Augmentation**

Data augmentation is a technique of artificially increasing the size and diversity of the training dataset by applying various transformations to the original images. Data augmentation can improve the performance and robustness of the deep learning algorithms, as it reduces the risk of overfitting and introduces more variability to the data. In this study, the following image augmentations will be used to enhance the underwater image dataset:

1. Rotation: randomly rotate the images by a range of 0.2 degrees, which can account for the orientation changes of the underwater objects.
2. Width and height shift: randomly shift the images horizontally and vertically by a range of 0.05, which can account for the position changes of the underwater objects.
3. Zoom: randomly zoom in or out the images by a range of 0.05, which can account for the scale changes of the underwater objects.
4. Horizontal flip: randomly flip the images horizontally, which can account for the symmetry of some underwater objects.

These augmentations will be applied to each image in the training and testing dataset with a probability of 0.5, resulting in a larger and more diverse dataset.

### **3.4 Image Segmentation**

Image segmentation is the process of dividing an image into regions that correspond to different objects or semantic categories. For

underwater object detection, image segmentation can help to identify and locate the objects of interest in the complex and noisy underwater environment. Image segmentation can also provide useful information for other tasks, such as image restoration, marine biology, and underwater robotics.

Learning-based semantic segmentation is a type of image segmentation that uses machine learning techniques, especially deep neural networks, to learn the mapping from the input image to the output segmentation map. Learning-based semantic segmentation can leverage a large amount of data and the powerful feature extraction capabilities of deep neural networks to achieve high accuracy and robustness. Learning-based semantic segmentation can be divided into two main categories: fully convolutional networks (FCNs) and encoder-decoder networks.

FCNs use convolutional layers to produce the segmentation map directly from the input image, without using any fully connected layers. Encoder-decoder networks use an encoder network to extract high-level features from the input image, and a decoder network to upsample the features and generate the segmentation map. Both types of networks can be trained end-to-end using supervised learning, where the ground truth segmentation maps are provided as labels. Learning-based semantic segmentation can handle various challenges of underwater images, such as low visibility, color distortion, and scale variation, by using appropriate network architectures and data augmentation techniques.



## **3.5 Image Segmentation using VGG-16 and Fully Convolutional Networks**

This section presents a method for image segmentation of underwater imagery, which is the task of assigning a label to each pixel in an image, such as background, foreground, or object classes. Image segmentation can be useful for many applications, such as scene understanding, object detection, medical image analysis, etc.

### **3.5.1 Network Architecture**

The proposed model is built upon a fully convolutional encoder-decoder architecture, which includes skip connections connecting corresponding layers. The base structure of the model adopts residual learning principles, with an additional optional skip layer named RSB (Residual Skip Block). Each RSB is composed of three convolutional (conv) layers, followed by Batch Normalization (BN) and ReLU non-linearity.

As illustrated in Figure 3.4, This approach is chosen to ensure real-time inference while maintaining a reasonable segmentation performance. Conversely, the model prioritizes improved performance, utilizing 12 encoding layers from a pre-trained VGG-16 architecture.

### **3.5.2 Training Pipeline and Implementation Details**

The approach to underwater object detection involves training a deep learning model to learn a mapping from the input domain of natural underwater images to their corresponding semantic labels in RGB space.

The standard cross-entropy loss is used as the supervisory signal for end-to-end training, which measures the difference between predicted and ground truth pixel labels. The optimization pipeline is implemented using TensorFlow libraries and trained on a Windows host with an NVIDIA GeForce GTX 1080 8GB graphics card. The Adam optimizer with a learning rate of  $10^{-4}$  and momentum of 0.5 is used for global iterative learning. To enhance the robustness of the model, various image transformations are applied for data augmentation during training.

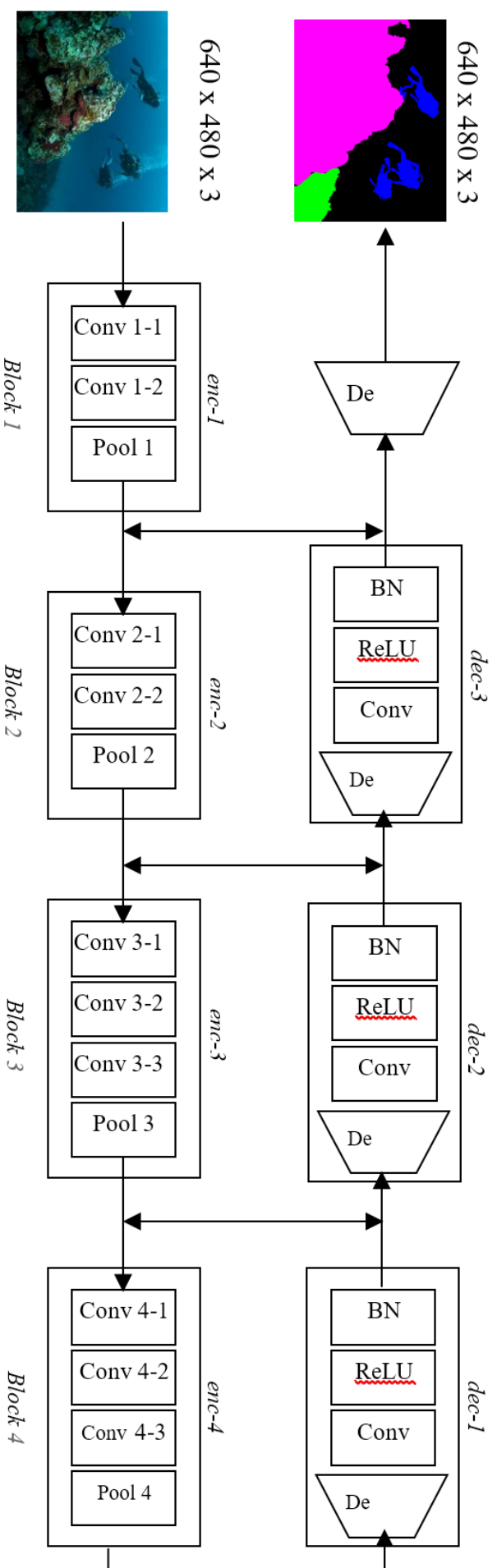



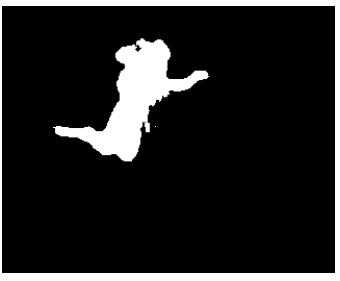

Figure 3.4 The Architecture of the proposed model which composed of the Initial Four Blocks from a Pretrained VGG-16 Model for Encodings.

This is Followed by Three Symmetrical Decoder Blocks and a Deconvolution Layer.

### 3.6 Post-Processing with Morphological Operations

Morphological operations are important techniques for underwater image processing, as they can shape and analyze the structures in images, improving their quality and interpretability. These operations use two basic processes: dilation and erosion. Dilation expands the boundaries of objects in an image, while erosion shrinks them. The amount of expansion or shrinkage depends on the size and shape of the structuring element, which is a small binary image that defines the neighborhood of a pixel.

One of the morphological operations that is used for underwater image processing is close dilation, which is a combination of dilation and erosion. Close dilation first dilates an image and then erodes the dilated image, using the same structuring element for both operations. Close dilation is useful for filling small holes in an image while preserving the shape and size of large holes and objects in the image<sup>1</sup>. Close dilation can also enhance the visibility and contrast of objects, which is crucial for underwater object detection. Shows an example of close dilation applied to an underwater image (see Figure 3.5).

Ground Truth	Before Morphological Operations	After Morphological Operations
		

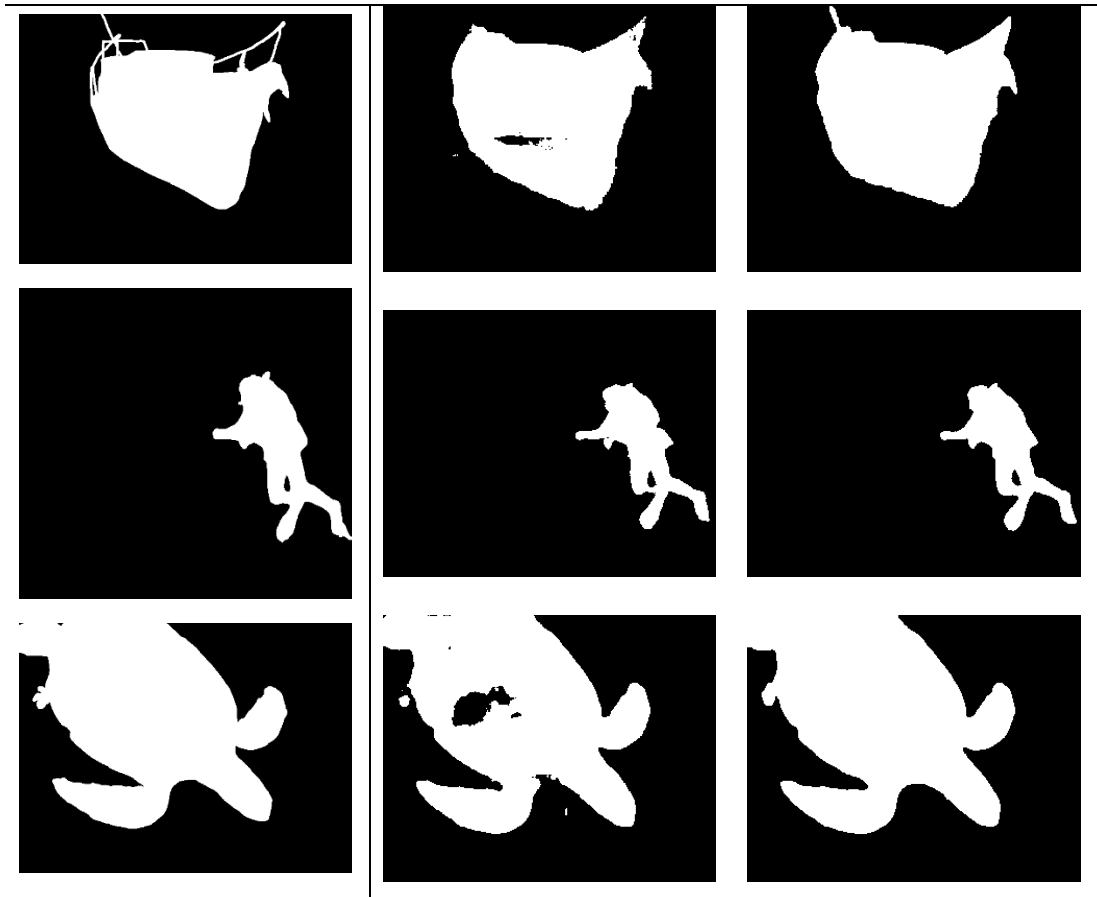


Figure 3.5 Enhanced Results Before and After Morphological Operations Application.

## **CHAPTER FOUR**

### **RESULTS AND DISCUSSION**

## 4.1 Overview

This chapter offers a comprehensive exploration and evaluation of the performance of the proposed model in the context of underwater object detection and segmentation. It encompasses a series of well-designed experiments, each aimed at assessing different aspects of the model's capabilities.

The main goal of this chapter is to present and analyze the quantitative results obtained from the model's performance across various benchmark metrics and experimental scenarios. These outcomes are systematically presented, providing insights into the strengths, limitations, and nuances of the model when compared to established state-of-the-art models. The chapter unfolds by providing a comprehensive overview of the methodologies, the validation dataset, and the chosen evaluation metrics. The execution of each experiment is documented, focusing on configuration settings, training procedures, and data augmentation techniques.

Following this, the chapter presents numerical results through tables, graphs, and visual representations, enabling a comparative analysis of the model's performance against benchmark models. Metrics related to object detection accuracy, semantic segmentation precision, and saliency prediction are thoroughly discussed in the context of the model's overall performance objectives.

As the chapter progresses, the discussion section takes a central role. The results are thoughtfully contextualized, interpreted, and compared with existing literature. Significant findings, trends, and implications emerging from the experimental outcomes are elucidated,

providing a comprehensive narrative on the importance and relevance of the model's contributions to the field of underwater object detection.

## **4.2 The Dataset**

The SUIM dataset[69] is a comprehensive collection of underwater images specifically designed for semantic segmentation tasks. It provides a valuable resource for advancing research and development in the field of underwater computer vision. With its diverse set of object categories and carefully annotated pixel-level labels, the dataset facilitates the training and evaluation of state-of-the-art semantic segmentation models for underwater scenes.

Underwater environments pose unique challenges for image analysis due to factors such as light attenuation, domain-specific object categories, background patterns, and optical distortion artifacts. These challenges necessitate the development of specialized algorithms and datasets tailored to the underwater domain. The SUIM dataset fills this gap by offering a large-scale annotated dataset that covers a wide range of object categories encountered in underwater exploration and surveying applications.

The SUIM dataset comprises several object categories that are labeled for semantic segmentation. These categories include: 1) Waterbody background (BW) 2) Human divers (HD) 3) Aquatic Plants/Flora (PF) 4) Wrecks/ruins (WR) 5) Robots and instruments (RO) 6) Reefs and other invertebrates (RI) 7) Fish and other vertebrates (FV) 8) Sea-floor and rocks (SR).

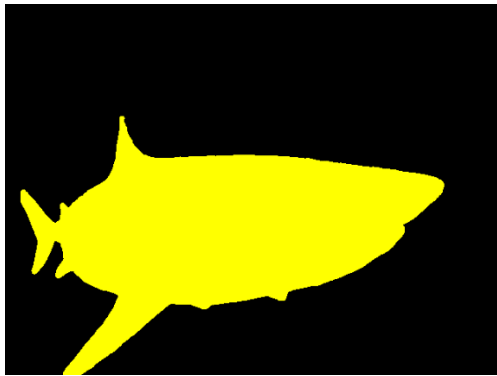


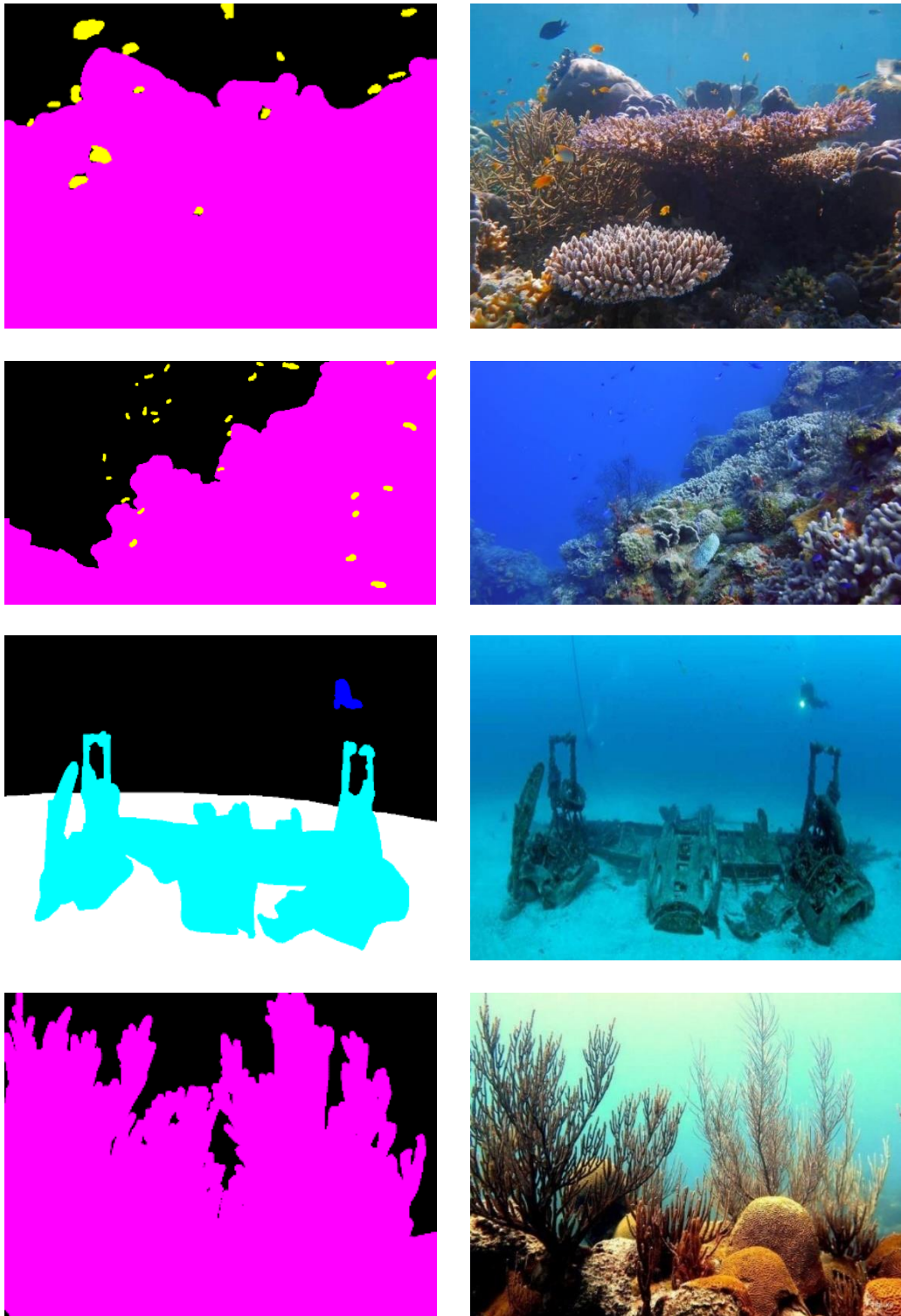
To represent these object categories in the image space, a 3-bit binary RGB color scheme is utilized. Each object category is assigned a specific color using this color representation. This color-based labeling approach enables the distinction and identification of different objects within the images.

These object categories and their corresponding colors are essential for training and evaluating semantic segmentation models on the SUIM dataset as shown in Table 4.1. They serve as the ground truth labels for the dataset (see Figure 4.1), allowing for an accurate assessment of model performance in detecting and segmenting the various underwater objects and features.

*Table 4.1 The Object Categories and Corresponding Color Codes for Pixel Annotations in The SUIM Dataset.*

<b>Object category</b>	<b>RGB Color</b>	<b>Code</b>
Background (waterbody)	000	BW
Human divers	001	HD
Aquatic plants and sea-grass	010	PF
Wrecks or ruins	011	WR
Robots (AUVs/ROVs/instruments)	100	RO
Reefs and invertebrates	101	RI
Fish and vertebrates	110	FV
Sea-floor and rocks	111	SR





*Figure 4.1 Sample of SUIM Dataset.*

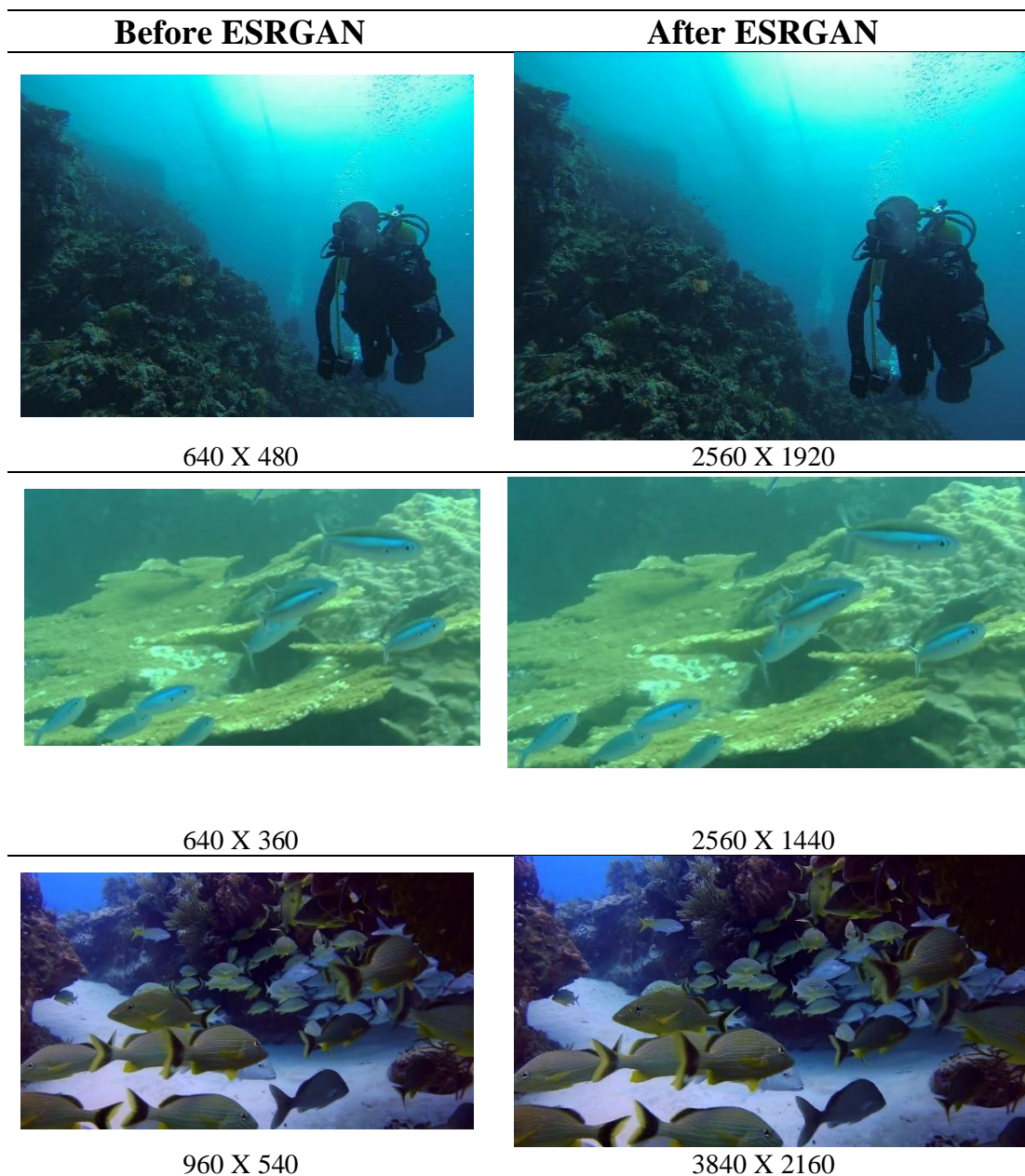
## **4.3 Pre-Processing**

Pre-processing is an important step in underwater image processing, as it can improve the quality and clarity of the images before applying further analysis and interpretation. Pre-processing techniques can address the challenges and limitations of underwater imaging, such as low resolution, noise, blurring, color distortion, and non-uniform illumination. This section presents the results of two pre-processing methods that were applied to the underwater image dataset: ESRGAN and image augmentation.

### **4.3.1 ESRGAN**

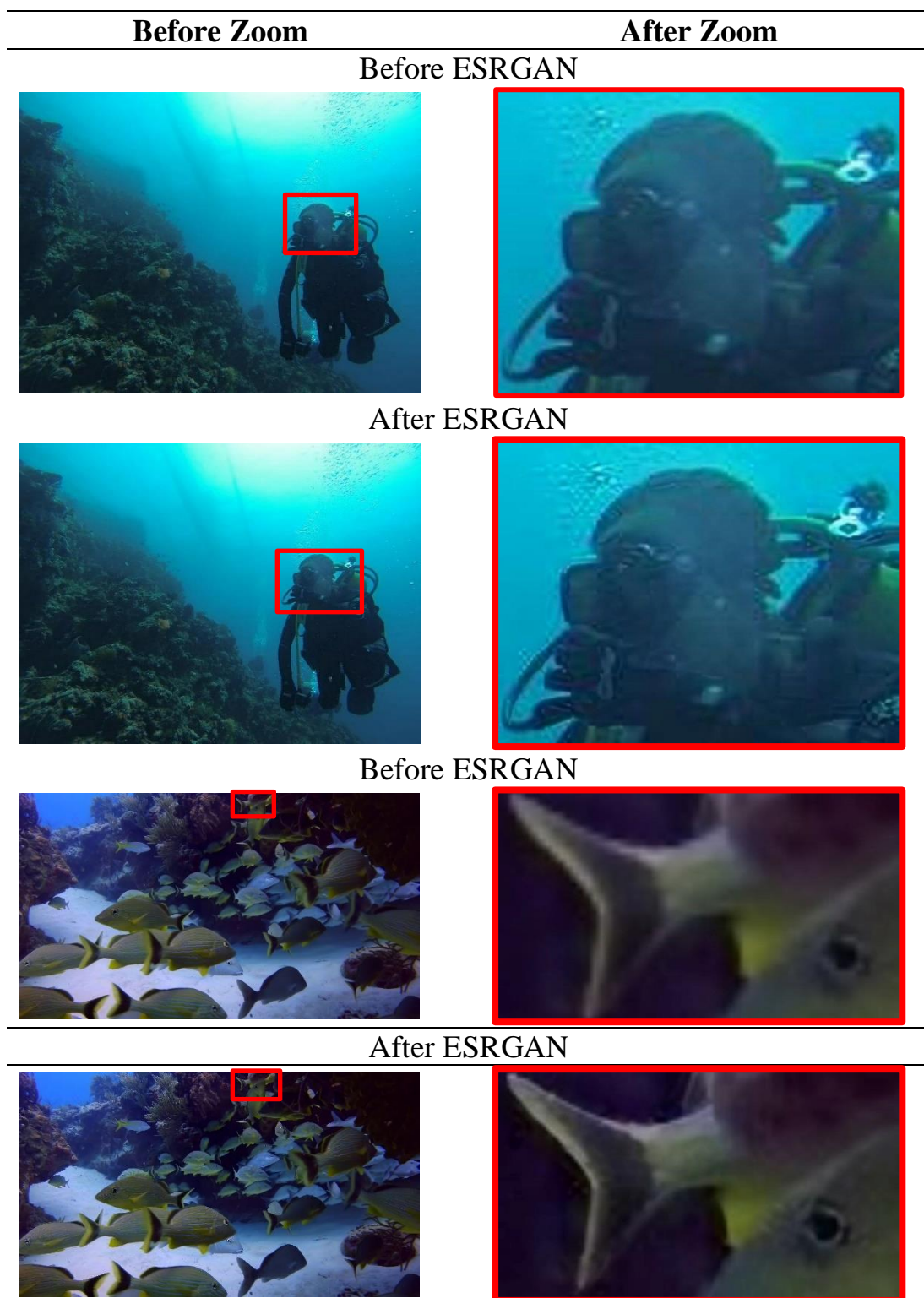
ESRGAN stands for Enhanced Super-Resolution Generative Adversarial Networks, which is a method for improving the quality and realism of images that are upscaled from low resolution to high resolution. ESRGAN is based on the Super-Resolution Generative Adversarial Network (SRGAN) but introduces several enhancements to the network architecture, the adversarial loss, and the perceptual loss. ESRGAN can produce high-resolution images that are sharper, more detailed, and more natural-looking.

The integration of the ESRGAN technique was an essential part of the exploration, as it contributed to the augmentation of image quality and, consequently, to the efficacy of the model's performance. The training phase started with a pre-trained model's parameter values as a starting point. The training conducted with ESRGAN showed promising results, as illustrated in Figure 4.2 and Figure 4.3.



*Figure 4.2 Visual Enhancement with ESRGAN: Sample Images from the SUIM Dataset Before and After Processing.*






*Figure 4.3 Visual Enhancement with ESRGAN - Zoomed-In Comparisons: Sample Images from the SUIM Dataset Before and After Processing to Highlight Effectiveness.*

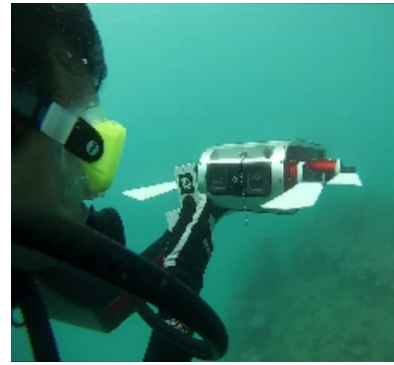
### 4.3.2 Dataset Image Augmentation

Image dataset augmentation involves applying various image processing operations, such as rotation, shifting in width and height, shearing, zooming, and horizontal flipping, to diversify the dataset. These operations introduce variations in the object's position and image quality, ultimately enhancing the dataset's diversity. This augmentation process aids in improving the training's generalization and consequently enhances the object detection capabilities. The specific parameters for these operations are determined by random values within a predefined range, as detailed in Table 4.2.

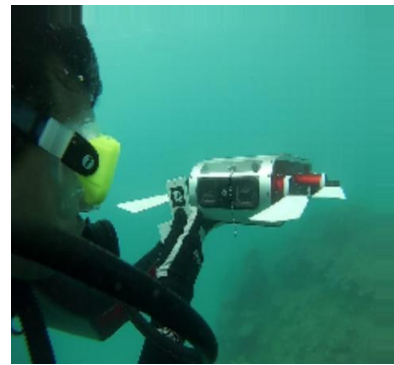
*Table 4.2 Image Processing Operations Configuration to Introduce Image Augmentation.*

<b>Image processing operation</b>	<b>Configuration bounds</b>
1. Original image	

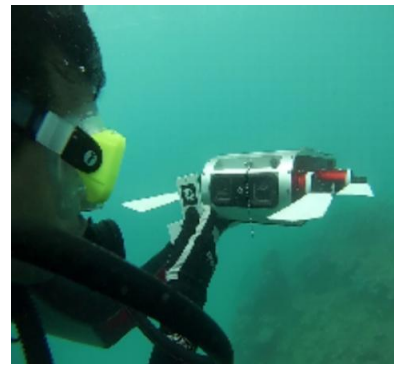
1. Rotation -0.2 to +0.2 radians



2. Width Shift -0.5 to +0.5 radians



3. Height Shift -0.5 to +0.5 radians



4. Shear -0.5 to +0.5 radians

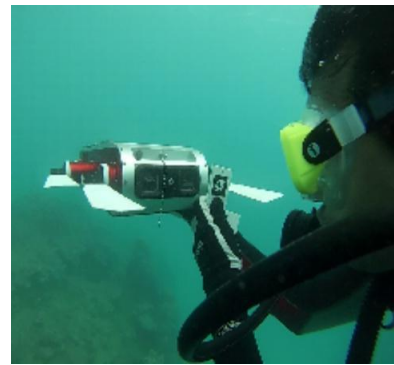




5. Zoom Range -0.5 to +0.5 radians



6. Horizontal flip enabled



---

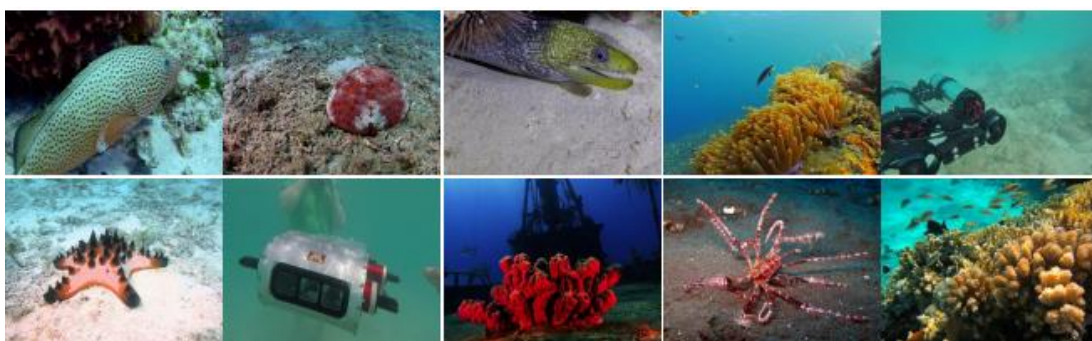
The SUIM dataset comprises a total of 1525 RGB images designated for training and validation purposes, along with an additional 110 images for testing to evaluate semantic segmentation models. These images exhibit varying spatial resolutions, ranging from dimensions like  $1906 \times 1080$ ,  $1280 \times 720$ ,  $640 \times 480$ , to  $256 \times 256$  pixels.

These images were meticulously selected from a vast pool of samples collected during underwater explorations and collaborative experiments involving humans and robots in diverse water environments. Some images were also drawn from larger datasets such as "EUVP" (Enhancing Underwater Visual Perception)[75] (see Figure 4.6), which represents a collection of techniques aimed at improving the visual quality of underwater images. In contrast, "UFO-120" (Underwater Florida Objects) UFO-120 [76] (see Figure 4.4) is a dataset focused on object detection and recognition in Florida's underwater conditions. "USR-248"

(Underwater Sonar and Radar)[50] (see Figure 4.5) is another dataset primarily used for underwater navigation and target detection through sonar and radar data. The "SUIM dataset" appears to be a specialized dataset tailored for the semantic segmentation of underwater imagery, allowing researchers to classify each pixel into predefined categories. These datasets and methods differ in terms of their type, purpose, and application, and each plays a unique role in advancing the field of underwater image analysis and enhancement. Which were previously designed for addressing underwater image enhancement and super-resolution challenges. The image selection process aimed to encompass a wide range of natural underwater scenes and different setups used in human-robot collaborative experiments.



*Figure 4.5 A Few Sample Images from The UFO-120 Dataset are Shown.*



*Figure 4.4 A Few Sample Images from The USR-248 Dataset are Shown.*



Figure 4.6 A few sample images from the EUVP dataset are shown.

To provide further insights into the dataset, Figure 4.7 illustrates the distribution of each object category, their relationships with one another, and the distribution of RGB channel intensity values within the SUIM dataset.

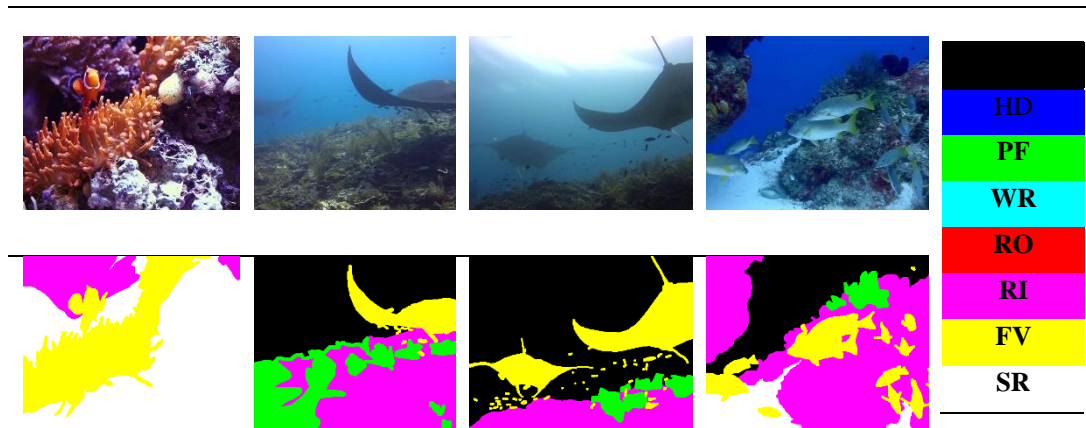
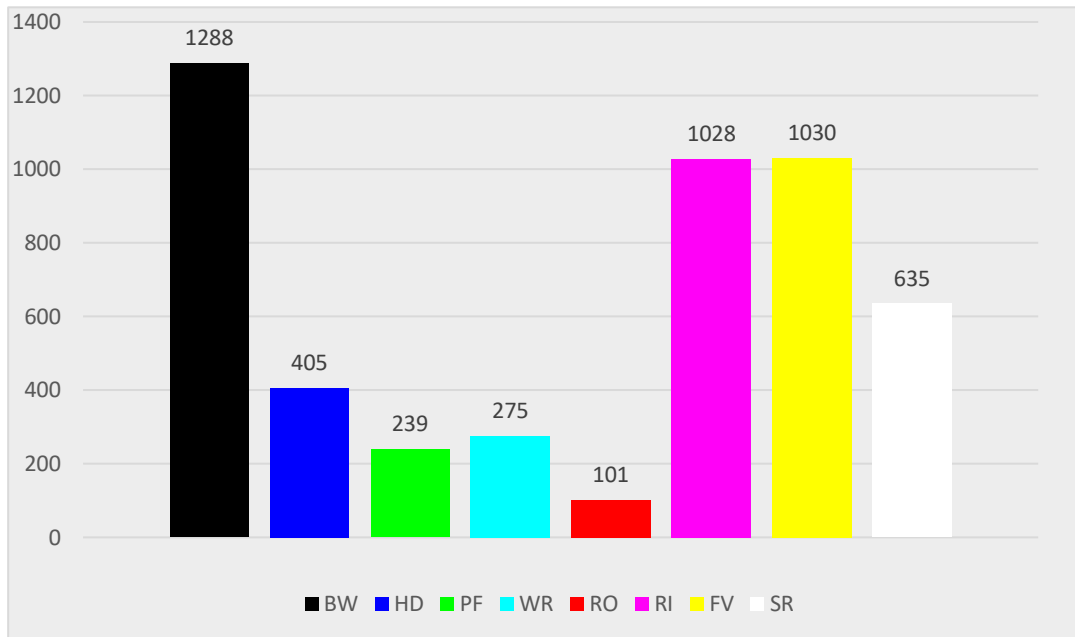


Figure 4.7 Sample Images and Corresponding Pixel Annotations in the SUIM Dataset.



*Figure 4.8 Statistics of Object Categories Values in the SUIM Dataset.*

## 4.4 Deep Learning Network Training Settings

As in the background chapter, it was demonstrated that semantic segmentation is an effective approach for object detection, particularly in the context of underwater environments. It offers a good mean average precision when combined with image super-resolution using ESRGAN.

The training settings for semantic segmentation are outlined in Table 4.3. These settings encompass the hardware utilized, training resolution, parameters to mitigate overfitting, learning rate, image augmentation techniques, saving of training parameters, and maximum iterations.

The training process was conducted on a GPU card, specifically the NVIDIA GeForce GTX 1080 with 8GB of memory. The training was performed over 5,000 iterations, employing various spatial resolutions,

such as  $1906 \times 1080$ ,  $1280 \times 720$ ,  $640 \times 480$ , and  $256 \times 256$  pixels. Network parameters were saved every 5,000 iterations to monitor the training progress.

To prevent overfitting, a set of parameters was employed. Simple image augmentation techniques were applied to the dataset, including a rotation range of 0.2, width shift, height shift, shear, and a zoom range of 0.05. Horizontal flip was enabled, while the remaining parameters were kept at their default values.

The learning rate, which influences how the optimizer adapts during training, was set to a dynamic value. A large learning rate can lead to rapid changes in weight values, potentially resulting in convergence to a suboptimal solution. Conversely, a low learning rate may cause slow convergence. Therefore, a dynamic learning rate was utilized to ensure a stable gradient and prevent model divergence.

By providing these detailed training settings, a robust framework is established for training the semantic segmentation model for underwater object detection. These settings aim to achieve accurate results and mitigate potential issues such as overfitting and unstable gradients.

*Table 4.3 Underwater Object Detection Training Settings.*

<b>Category</b>	<b>Configuration item</b>	<b>Configuration value</b>
1. Network	Deep learning network	CNN
2. Hardware	GPU card used	Nvidia GTX 1080
3. Training Resolution	Image resolution during training	$1906 \times 1080$ , $1280 \times 720$ , $640 \times 480$ , and $256 \times 256$ pixels
4. Learning Rate Adjustment	Learning rate	0.0001

5. Image Augmentations	rotation range	0.2
	Width shift range	0.05
	Height shift range	0.05
	Zoom range	0.05
	Horizontal flip	enabled
6. Epoch Number	Number of Epochs	50
7. Data Saving	Save data every	5,000 Iteration
8. Total Training Iterations	Total Iterations	250,000 Iteration

---

## 4.5 Underwater Object Detection Using Deep Learning Model Training and Results

In this thesis, a specialized underwater object detection model materialized through meticulous development and configuration. The bedrock of its construction rested upon the precise parameters and environmental settings, thoughtfully outlined in section 4.1. This model undertook rigorous training, immersing itself in a dataset of underwater images as meticulously elucidated in section 4.2.

A distinctive facet of this study involved training the model with unwavering attention to meticulous detail. Before integrating ESRGAN as a preprocessing step, training each epoch took roughly 10 hours, guided by the parameters detailed in Table 1.

However, following the introduction of ESRGAN to enhance image quality, the training time per epoch was extended to nearly 12 hours due to increased computational demand. Over 50 epochs, this resulted in an extensive 25-day training duration. This illustrates the trade-off between improved image quality through ESRGAN and the extended temporal commitment for training in underwater image analysis.

Remarkably, this intensive training was conducted on a curated dataset comprising 1,525 images, each intricately capturing the essence of underwater scenarios.

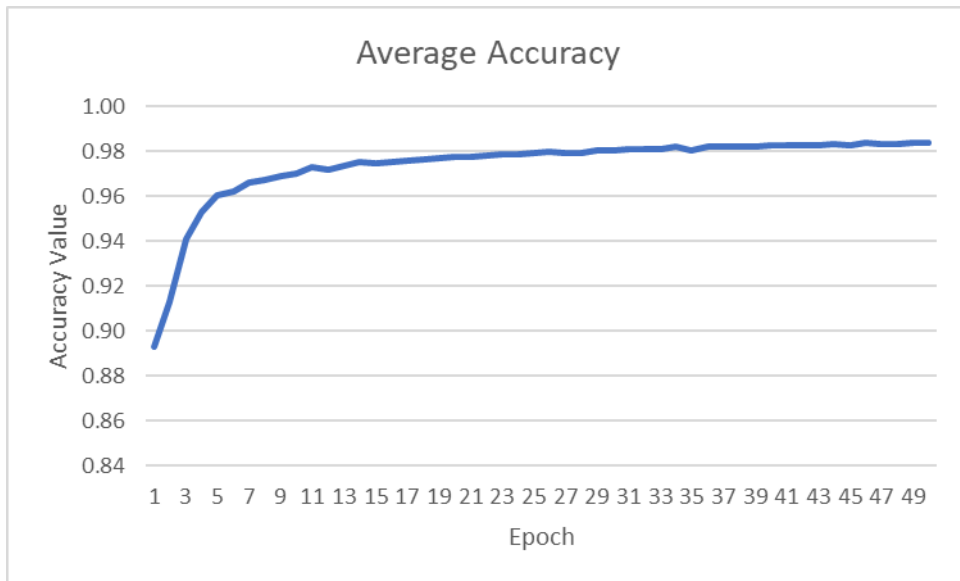
The imminent presentation, a beacon of forthcoming insights, promises to unveil an all-encompassing perspective encompassing the training dynamics and the consequential evaluation metrics for other model iterations. This calculated juxtaposition seeks to offer a comprehensive vantage point into the diverse performance of these models in the intricate task of detecting submerged objects within underwater imagery. The spectrum of evaluation metrics spans a diverse range of object detection indices, with particular emphasis on pivotal measures such as Average Precision (AP), mean Intersection over Union (mIoU), precision, recall, and the F1-score. Additionally, it is essential to underscore that the exploration encompassed the integration of the ESRGAN technique. This technique's integration presents an additional dimension, contributing to the augmentation of image quality and, consequently, to the efficacy of the model's performance.

Initiating the training phase, the model leverages a pre-trained model's parameter values as a starting point. Subsequently, the assessment of outcomes commences through the evaluation of the loss function. Typically, this process begins with a notably high loss, gradually diminishing as optimization of the model's parameters unfolds. As evidenced in the training conducted with ESRGAN, the initial average accuracy of 0.8932 ascended commendably to 0.9836, a trend showcased in Figure 4.9. Simultaneously, the average loss, commencing at 0.0713, exhibited a downward, culminating at 0.0101, as illustrated in Figure 4.10.

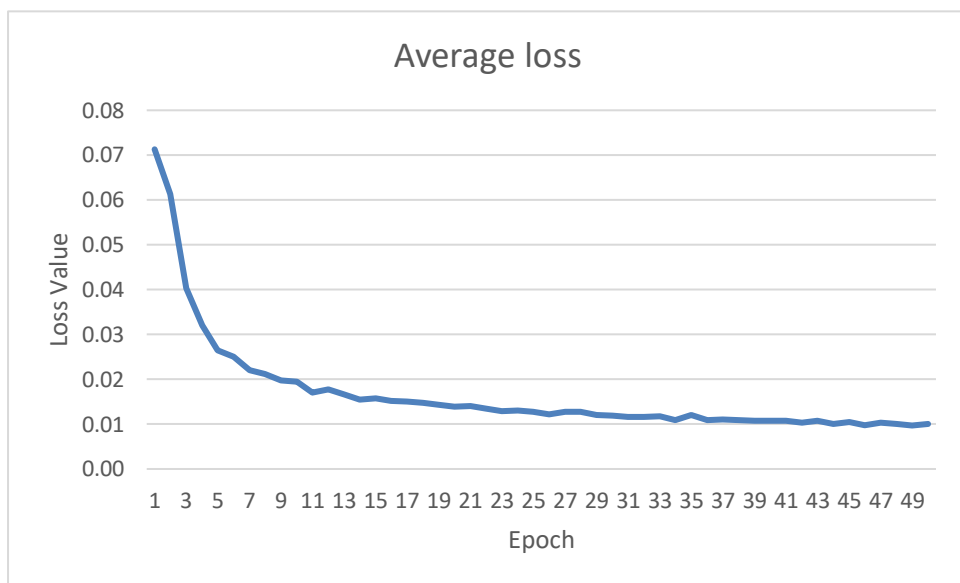
These outcomes affirm the model's adaptability to the underlying challenge.

However, the optimizer's proficiency in problem adaptation and resolution does not inherently guarantee the trained model's precision in object detection. Diverse factors may contribute to this, ranging from an insufficiently diverse training dataset, which may fail to encapsulate all conceivable object scenarios, to potential limitations rooted in the quality of the image dataset itself. Challenges might arise from inefficiencies in extracting object features due to inadequate image quality, further accentuating the complexity of the detection process.





*Figure 4.10 Accuracy versus different epoch plot.*

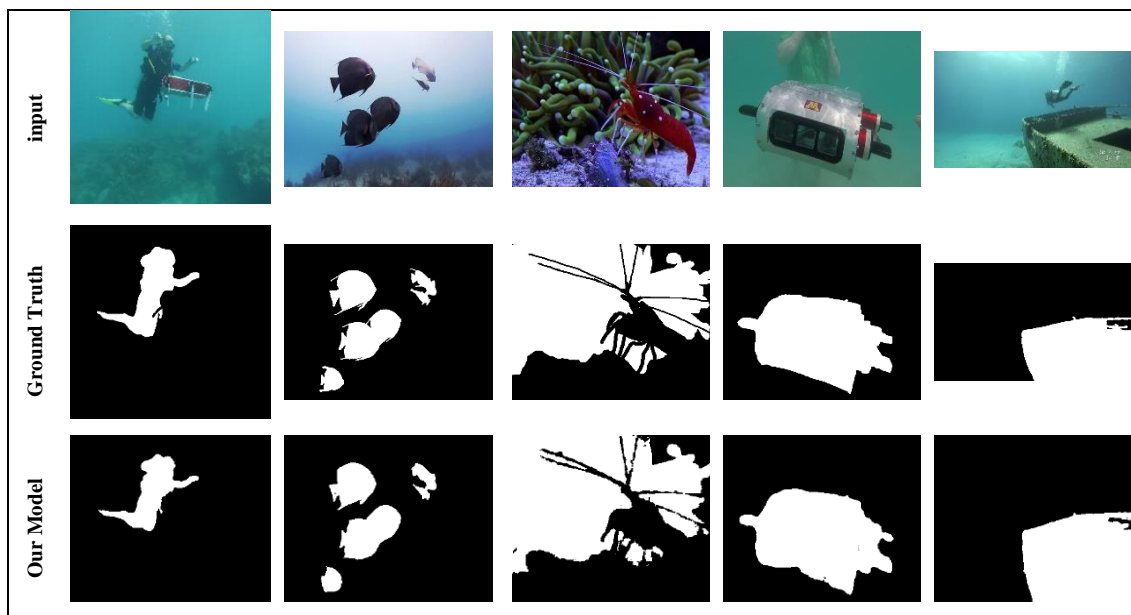


*Figure 4.9 Loss versus different epoch plot.*

In Section 3, the SUIM dataset and its diverse applications in semantic segmentation and saliency prediction were thoroughly explored. In the evaluation, the performance of state-of-the-art (SOTA) models is compared using the following two training configurations:
















### 1. Semantic Segmentation with Five Major Object Categories

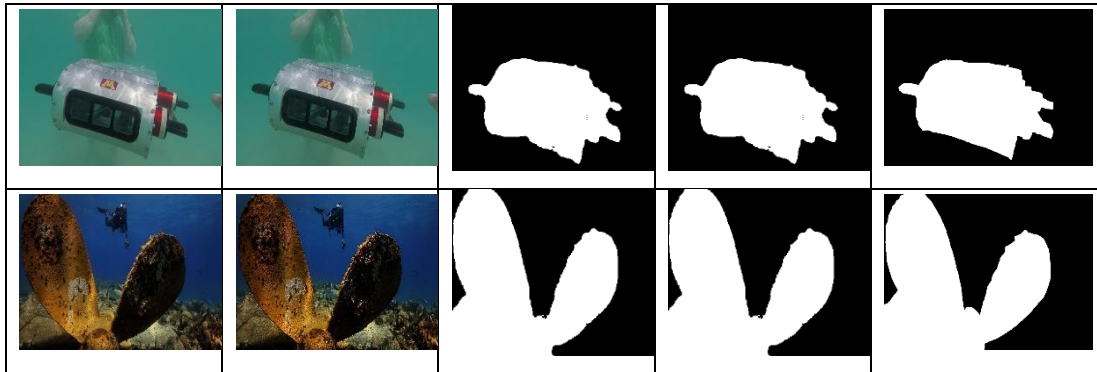
This configuration involves semantic segmentation of images into five major object categories, as detailed in Table 4.2: HD, WR, RO, RI, and FV. All other objects are considered as background, denoted as BW=PF=SR= (000)RGB. Each model is set up to produce five channels of output, one for each of these object categories. The separate pixel masks predicted by the models are then combined to create RGB masks for visualization(see Figure 4.11).



*Figure 4.11 Qualitative Comparison of Semantic Segmentation with Object Categories (HD, WR, RO, RI, and FV) Using the Proposed Model and Ground Truth.*

Figure 4.12 presents a set of image samples, showcasing the various stages of the image processing pipeline. The first column displays the original images, while the second column reveals the images enhanced using the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network). Moving to the third and fourth columns, images before and after applying morphological operations are observed, which play a pivotal role in refining the segmentation results. Finally, the fifth column features the ground truth images, providing a reference for evaluating the quality of the image processing techniques. These image samples illustrate the progressive improvements achieved through each stage of the processing pipeline, emphasizing the significance of ESRGAN and morphological operations in enhancing image quality and facilitating more accurate object detection and segmentation.

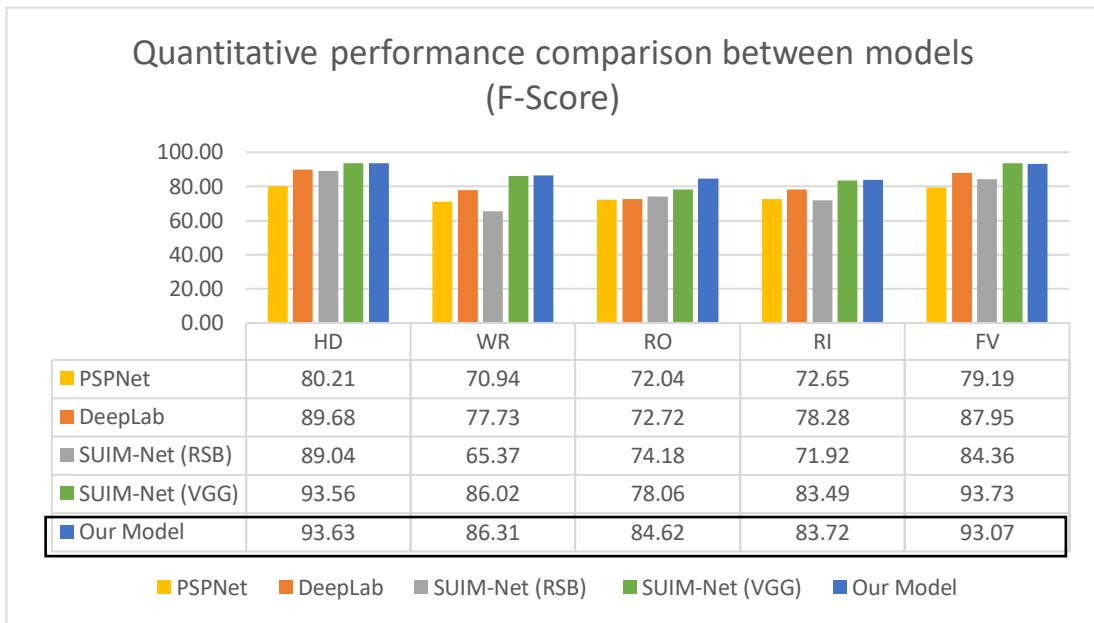
Original	ESRGAN	Before Morphological	After Morphological	Ground Truth
				
				
				



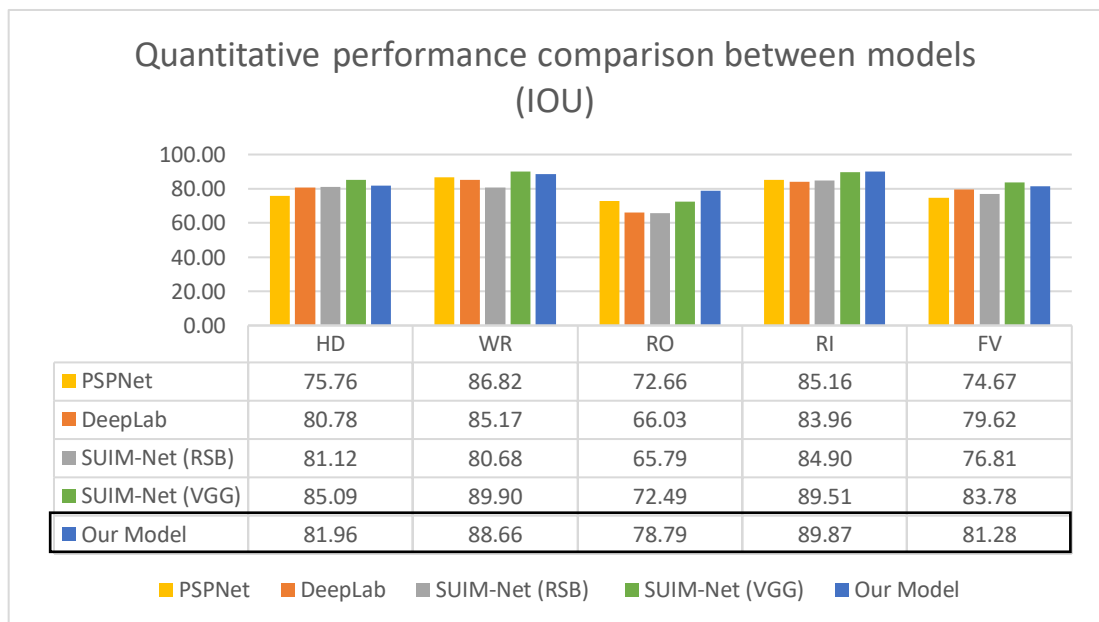
*Figure 4.12 Image Samples — Original, ESRGAN-Enhanced, Pre-Morphological, Post-Morphological, and Ground Truth Images.*

## **2. Single-Channel Saliency Prediction**

In this configuration, the ground truth intensities of HD, RO, FV, and WR pixels are assigned a value of 1.0, while the rest are set to 0.0. The model's output is thresholded and visualized as binary images for saliency prediction. In the comprehensive evaluation, the performance of all models was meticulously assessed using established metrics for region similarity and contour accuracy.



*Figure 4.14 Quantitative performance comparison between models to show the F-Score.*



*Figure 4.13 Quantitative performance comparison between models to show the IOU.*

Quantitative outcomes, as depicted in Figures 4.13 and 4.14, offer a comparative analysis of F-Score and mean IOU (mIOU) scores, spanning each object class for semantic segmentation and saliency

prediction tasks. Amid the evaluated models, DeepLabV3 consistently emerged as the frontrunner, boasting the three highest F-Score and mIOU scores for both tasks. Notably, The combination of PSPNet with MobileNet, often denoted as PSPNet<sub>MobileNet</sub>, harnesses the strengths of both architectures. The MobileNet backbone provides efficiency, making it feasible for real-time or mobile deployment, while the PSPNet module enhances its semantic segmentation capabilities by capturing scene context at multiple scales.

Conversely, the SUIM-Net<sub>RSB</sub> and SUIM-Net<sub>VGG</sub> models [69] exhibited unwavering and competitive performance, evident in terms of region similarity and object localization. This resilience was reflected in the minor deviations from the respective highest scores.

Figure 4.15 visually encapsulates the average F-Score and IOU comparison, showcasing the superiority of the proposed model in terms of accuracy and object boundary localization for underwater semantic

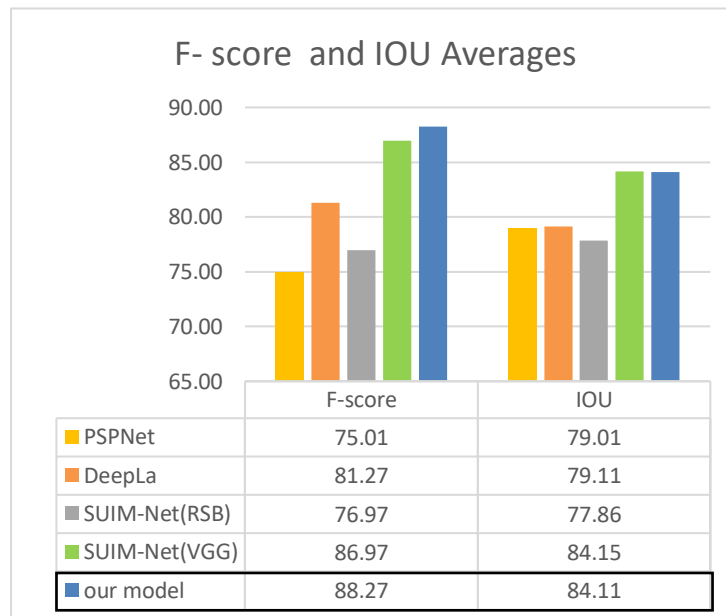


Figure 4.15 The Average of F-Score and IOU.

segmentation. Intriguingly, the proposed model demonstrated notable enhancements in accuracy for specific objects, as illustrated alongside. These advancements further underscore the efficacy of the approach and its potential for superior performance in semantic segmentation and saliency prediction tasks compared to the other evaluated models.

Table 4.4 presents a comparative analysis of F-measure scores across various object detection methods, focusing on the impact of Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) and Morphological Processing. The table contrasts F-measure scores before and after applying these enhancement techniques, shedding light on their effectiveness in improving object detection accuracy. Additionally, the scores are juxtaposed with a State-of-the-Art (SOTA) Benchmark to assess the relative performance and advancements achieved. This comparative assessment is instrumental in evaluating the practical benefits of employing ESRGAN and Morphological Processing in underwater object detection applications.

*Table 4.4 Comparison of F-measure Scores Before and After ESRGAN and Morphological Processing with SOTA Benchmark.*

<b>F-measure</b>				
Before ESRGAN	After ESRGAN	Before Morphological	After Morphological	SOTA
86.41	87.1426	87.53	88.27	86.97

## **CHAPTER FIVE**

# **CONCLUSION AND FUTURE WORKS**



## **5.1 Overview**

In this chapter, a comprehensive exploration of underwater object detection using deep learning techniques has been presented. The research aimed to address the challenges posed by the unique characteristics of underwater environments, such as poor visibility, varying lighting conditions, and complex backgrounds. Through extensive experimentation and analysis, promising results and insights have been achieved, although several avenues for improvement and further research remain.

## **5.2 Conclusion**

In conclusion, this research has addressed a critical need in the field of underwater computer vision by providing robust solutions for semantic segmentation and pixel-level detection of salient objects, enhancing the capabilities of visually-guided Autonomous Underwater Vehicles (AUVs). The main contributions and implications of this research are:

- 1- The SUIM dataset was introduced as a significant milestone in the domain of underwater computer vision. This extensively annotated dataset was created to perform general semantic segmentation of underwater environments, covering five distinct object categories: human divers (HD), fish and vertebrates (FV), robots (RO), reefs and invertebrates (RI), and wrecks and ruins (WR).
- 2- VGG-16, a fully-convolutional encoder-decoder architecture, was developed, that outperforms existing state-of-the-art models in

terms of both semantic segmentation accuracy and computational efficiency. With an 88% accuracy rate, this model excels in accurately detecting and classifying objects in challenging underwater environments.

- 3- Image Super-Resolution using ESRGAN was employed as a preprocessing step to enhance the resolution and quality of low-resolution underwater images. Additionally, morphological operations were employed to further refine the segmentation results.

This research not only addresses current limitations but also paves the way for enhanced capabilities and practical use of visually-guided AUVs in underwater exploration, marine research, and environmental monitoring. By contributing to the advancement of underwater robotics and exploration, this work has the potential to significantly impact the understanding and preservation of underwater ecosystems and marine resources, making a valuable contribution to the field of marine science and technology.

### **5.3 Future Work**

This thesis introduces novel solutions for underwater image enhancement and underwater object detection. The proposed frameworks are designed to enhance and detect objects robustly in underwater environments, with extensive experimental results showcasing their high

accuracy and potential utility for ocean scientists and biologists. However, there remains room for improvement in the future endeavors.

- 1- The next step involves formulating a comprehensive enhancement-detection framework that seamlessly integrates low-level enhancement and high-level detection within a single end-to-end structure. The current approach employs distinct enhancement and detection models, leading to a sequential processing pipeline where enhanced images are sequentially fed into the detection model. This results in added time and resource overhead due to image transmission. In future work, the focus will be on developing a unified enhancement-detection framework that obviates the need for separate processes. Such a unified framework offers two advantages: firstly, joint optimization of both tasks is expected to yield better solutions compared to the cascaded pipeline; secondly, this integration is expected to significantly expedite the processing speed by eliminating the need for inter-model image transmission. The aim is to devise a novel multi-task loss function to evaluate the performance of this unified framework.
- 2- The demand for reduced computational complexity is critical for real-time applications in underwater scenarios. While a powerful deep detection framework was introduced in Chapter 4 that effectively addresses noise challenges in underwater data, its time complexity remains high due to its deep ensemble nature. Additionally, the considerable storage and memory bandwidth consumed by deep neural networks raises concerns. To address these issues, future research will incorporate various model compression algorithms into the

framework, aiming to substantially reduce memory and computational overhead. These algorithms, involving weight binarization, weight pruning, and compact block design, can lead to efficient training and inference in deep networks. The goal is to introduce a deep-learning compression algorithm that minimizes storage and energy requirements, rendering deep networks suitable for real-time deployment on AUVs and ROVs.

- 3- The plan for the future involves extending the enhancement-detection framework to accommodate underwater video analysis. For the analysis of sequential underwater data, RNNs are particularly well-suited. Unlike CNNs, RNNs feature self-connected hidden layers that facilitate information feedback across sequential inputs. This memory retention capacity allows RNNs to process varying sequences effectively. Hence, the strategy entails amalgamating CNNs and RNNs-based architectures into a unified deep-learning framework, thereby enabling comprehensive end-to-end analysis of underwater videos.

Furthermore, in future research endeavors, the incorporation of the YOLO framework is contemplated. YOLO is a widely recognized object detection architecture known for its real-time performance and accuracy. By integrating YOLO into the proposed enhancement-detection framework, even greater efficiency and precision in detecting objects within underwater scenes are anticipated. This strategic integration holds promise for enhancing the overall robustness and applicability of the model, thereby contributing to the advancement of underwater object detection and analysis.

## **REFERENCES**

## REFERENCES

- [1] M. Moniruzzaman, S. M. S. Islam, M. Bennamoun, and P. Lavery, "Deep learning on underwater marine object detection: A survey," in *Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18*, 2017, pp. 150-160: Springer.
- [2] Y. Zhou, Q. Wu, K. Yan, L. Feng, W. J. I. T. o. C. Xiang, and S. f. V. Technology, "Underwater image restoration using color-line model," vol. 29, no. 3, pp. 907-911, 2018.
- [3] D. Mallet and D. J. F. R. Pelletier, "Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012)," vol. 154, pp. 44-62, 2014.
- [4] A. Sahoo, S. K. Dwivedy, and P. J. O. E. Robi, "Advancements in the field of autonomous underwater vehicle," vol. 181, pp. 145-160, 2019.
- [5] I. Carlucho, M. De Paula, S. Wang, Y. Petillot, G. G. J. R. Acosta, and A. Systems, "Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning," vol. 107, pp. 71-86, 2018.
- [6] L. Chen *et al.*, "SWIPENET: Object detection in noisy underwater images," 2020.
- [7] P. I. Macreadie *et al.*, "Eyes in the sea: unlocking the mysteries of the ocean using industrial, remotely operated vehicles (ROVs)," vol. 634, pp. 1077-1091, 2018.
- [8] X. Ye *et al.*, "Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks," vol. 30, no. 11, pp. 3995-4008, 2019.
- [9] C. Li, S. Anwar, and F. J. P. R. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," vol. 98, p. 107038, 2020.
- [10] J. Y. Chiang and Y.-C. J. I. t. o. i. p. Chen, "Underwater image enhancement by wavelength compensation and dehazing," vol. 21, no. 4, pp. 1756-1769, 2011.
- [11] Z. Wang *et al.*, "UDD: An underwater open-sea farm object detection dataset for underwater robot picking," p. arXiv: 2003.01446, 2020.
- [12] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, and D. Kriegman, "Automated annotation of coral reef survey images," in *2012 IEEE conference on computer vision and pattern recognition*, 2012, pp. 1170-1177: IEEE.
- [13] K. Sarma and P. Vigneshwaran, "Underwater image enhancement using deep learning," in *Second International Conference on Image Processing and Capsule Networks: ICIPCN 2021 2*, 2022, pp. 431-445: Springer.
- [14] S. Gangisetty and R. R. J. S. P. I. C. Rai, "FloodNet: Underwater image restoration based on residual dense learning," vol. 104, p. 116647, 2022.
- [15] T. Li, Q. Yang, S. Rong, L. Chen, and B. J. A. O. He, "Distorted underwater image reconstruction for an autonomous underwater vehicle based on a self-attention generative adversarial network," vol. 59, no. 32, pp. 10049-10060, 2020.

- [16] R. Schettini and S. J. E. j. o. a. i. s. p. Corchs, "Underwater image processing: state of the art of restoration and image enhancement methods," vol. 2010, pp. 1-14, 2010.
- [17] J. Wu, X. Liu, Q. Lu, Z. Lin, N. Qin, and Q. J. S. P. I. C. Shi, "FW-GAN: Underwater image enhancement using generative adversarial network with multi-scale fusion," vol. 109, p. 116855, 2022.
- [18] S. Zhang, T. Wang, J. Dong, and H. J. N. Yu, "Underwater image enhancement via extended multi-scale Retinex," vol. 245, pp. 1-9, 2017.
- [19] J. Y. Chiang, Y.-C. Chen, and Y.-F. Chen, "Underwater image enhancement: using wavelength compensation and image dehazing (WCID)," in *Advanced Concepts for Intelligent Vision Systems: 13th International Conference, ACIVS 2011, Ghent, Belgium, August 22-25, 2011. Proceedings 13*, 2011, pp. 372-383: Springer.
- [20] D. Berman, D. Levy, S. Avidan, T. J. I. t. o. p. a. Treibitz, and m. intelligence, "Underwater single image color restoration using haze-lines and a new quantitative dataset," vol. 43, no. 8, pp. 2822-2837, 2020.
- [21] Z. Wang *et al.*, "Towards fairness in visual recognition: Effective strategies for bias mitigation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8919-8928.
- [22] Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong, and S. X. Yu, "Large-scale long-tailed recognition in an open world," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2537-2546.
- [23] C. Liu *et al.*, "A new dataset, Poisson GAN and AquaNet for underwater object grabbing," vol. 32, no. 5, pp. 2831-2844, 2021.
- [24] Y. Song, D. Nakath, M. She, K. J. P. J. o. P. Köser, Remote Sensing, and G. Science, "Optical imaging and image restoration techniques for deep ocean mapping: a comprehensive survey," vol. 90, no. 3, pp. 243-267, 2022.
- [25] N. bt Shamsuddin, B. b Baharudin, M. Kushairi, M. Rajuddin, and F. bt Mohd, "Significance level of image enhancement techniques for underwater images," in *2012 International Conference on Computer & Information Science (ICIS)*, 2012, vol. 1, pp. 490-494: IEEE.
- [26] R. Boluda-Ruiz, P. Rico-Pinazo, B. Castillo-Vázquez, A. García-Zambrana, and K. J. I. P. J. Qaraq, "Impulse response modeling of underwater optical scattering channels for wireless communication," vol. 12, no. 4, pp. 1-14, 2020.
- [27] S. Anwar, C. Li, and F. J. a. p. a. Porikli, "Deep underwater image enhancement," 2018.
- [28] N. G. Jerlov, *Marine optics*. Elsevier, 1976.
- [29] A. S. A. Ghani and N. A. M. J. A. s. c. Isa, "Underwater image quality enhancement through integrated color model with Rayleigh distribution," vol. 27, pp. 219-230, 2015.
- [30] I. Patel, S. J. J. J. o. A. S. Patel, and Computations, "Analysis of Various Image Segmentation Techniques for Flower Images," vol. 6, pp. 1936-1943.
- [31] Y. Yang, S. Hallman, D. Ramanan, C. C. J. I. T. o. P. A. Fowlkes, and M. Intelligence, "Layered object models for image segmentation," vol. 34, no. 9, pp. 1731-1743, 2011.

- [32] Y. Yu *et al.*, "Techniques and challenges of image segmentation: A review," vol. 12, no. 5, p. 1199, 2023.
- [33] D. Kaur, Y. J. I. J. o. C. S. Kaur, and M. Computing, "Various image segmentation techniques: a review," vol. 3, no. 5, pp. 809-814, 2014.
- [34] J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: an overview," in *Journal of physics: conference series*, 2018, vol. 1142, p. 012012: IOP Publishing.
- [35] J. Bell, *Machine learning: hands-on for developers and technical professionals*. John Wiley & Sons, 2020.
- [36] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [37] K. K. Verma, B. M. Singh, and A. J. I. J. o. I. T. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system," pp. 1-14, 2019.
- [38] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, A. J. J. S. Aljaaf, and u. l. f. d. science, "A systematic review on supervised and unsupervised machine learning algorithms for data science," pp. 3-21, 2020.
- [39] R. Rojas, *Neural networks: a systematic introduction*. Springer Science & Business Media, 2013.
- [40] L. N. Long, A. J. J. o. A. C. Gupta, Information,, and Communication, "Scalable massively parallel artificial neural networks," vol. 5, no. 1, pp. 3-15, 2008.
- [41] S.-H. Han, K. W. Kim, S. Kim, Y. C. J. D. Youn, and N. Disorders, "Artificial neural network: understanding the basic concepts without mathematics," vol. 17, no. 3, pp. 83-89, 2018.
- [42] K. Gurney, *An introduction to neural networks*. CRC press, 2018.
- [43] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object detection using convolutional neural networks," in *TENCON 2018-2018 IEEE Region 10 Conference*, 2018, pp. 2023-2027: IEEE.
- [44] Y. LeCun, Y. Bengio, and G. J. n. Hinton, "Deep learning," vol. 521, no. 7553, pp. 436-444, 2015.
- [45] J. D. Kelleher, *Deep learning*. MIT press, 2019.
- [46] A. K. Gupta, A. Seal, M. Prasad, and P. J. E. Khanna, "Salient object detection techniques in computer vision—A survey," vol. 22, no. 10, p. 1174, 2020.
- [47] N. Chen, W. Liu, R. Bai, and A. J. A. I. R. Chen, "Application of computational intelligence technologies in emergency management: a literature review," vol. 52, pp. 2131-2168, 2019.
- [48] H. Qin, X. Li, J. Liang, Y. Peng, and C. J. N. Zhang, "DeepFish: Accurate underwater live fish recognition with a deep architecture," vol. 187, pp. 49-58, 2016.
- [49] H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, and A.-Y. J. N. Zang, "Faster R-CNN for marine organisms detection and recognition using data augmentation," vol. 337, pp. 372-384, 2019.



- [50] M. J. Islam, S. S. Enan, P. Luo, and J. Sattar, "Underwater image super-resolution using deep residual multipliers," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 900-906: IEEE.
- [51] X. Wang *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0-0.
- [52] A. J. a. p. a. Aghelan, "Underwater Images Super-Resolution Using Generative Adversarial Network-based Model," 2022.
- [53] H. Wang *et al.*, "Simultaneous restoration and super-resolution GAN for underwater image enhancement," vol. 10, p. 1162295, 2023.
- [54] J. Song, H. Yi, W. Xu, X. Li, B. Li, and Y. J. H. Liu, "ESRGAN-DP: Enhanced super-resolution generative adversarial network with adaptive dual perceptual loss," vol. 9, no. 4, 2023.
- [55] S. Albahli, N. Nida, A. Irtaza, M. H. Yousaf, and M. T. J. I. a. Mahmood, "Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour," vol. 8, pp. 198403-198414, 2020.
- [56] M.-X. He, P. Hao, and Y.-Z. J. I. A. Xin, "A robust method for wheatear detection using UAV in natural scenes," vol. 8, pp. 189043-189053, 2020.
- [57] X. Wang, G. Xue, S. Huang, Y. J. J. o. M. S. Liu, and Engineering, "Underwater Object Detection Algorithm Based on Adding Channel and Spatial Fusion Attention Mechanism," vol. 11, no. 6, p. 1116, 2023.
- [58] D. Kim, D. Lee, H. Myung, and H.-T. J. I. S. R. Choi, "Artificial landmark-based underwater localization for AUVs using weighted template matching," vol. 7, pp. 175-184, 2014.
- [59] M.-C. Chuang, J.-N. Hwang, and K. J. I. T. o. I. P. Williams, "A feature learning and object recognition framework for underwater fish images," vol. 25, no. 4, pp. 1862-1872, 2016.
- [60] K. Blanc, D. Lingrand, and F. Precioso, "Fish species recognition from video using SVM classifier," in *Proceedings of the 3rd ACM International Workshop on Multimedia Analysis for Ecological Data*, 2014, pp. 1-6.
- [61] S. Villon, M. Chaumont, G. Subsol, S. Villéger, T. Claverie, and D. Mouillot, "Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between Deep Learning and HOG+ SVM methods," in *Advanced Concepts for Intelligent Vision Systems: 17th International Conference, ACIVS 2016, Lecce, Italy, October 24-27, 2016, Proceedings 17*, 2016, pp. 160-171: Springer.
- [62] A. Rova, G. Mori, and L. M. Dill, "One fish, two fish, butterflyfish, trumpeter: Recognizing fish in underwater video," in *MVA*, 2007, pp. 404-407.
- [63] A. Krizhevsky, I. Sutskever, and G. E. J. C. o. t. A. Hinton, "Imagenet classification with deep convolutional neural networks," vol. 60, no. 6, pp. 84-90, 2017.
- [64] S. Choi, "Fish Identification in Underwater Video with Deep Convolutional Neural Network: SNUMedinfo at LifeCLEF Fish task 2015," in *CLEF (Working Notes)*, 2015, pp. 1-10.

- [65] X. Li, M. Shang, H. Qin, and L. Chen, "Fast accurate fish detection and recognition of underwater images with fast r-cnn," in *OCEANS 2015-MTS/IEEE Washington*, 2015, pp. 1-5: IEEE.
- [66] S. Ren, K. He, R. Girshick, and J. J. A. i. n. i. p. s. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," vol. 28, 2015.
- [67] H. Yang, P. Liu, Y. Hu, and J. J. M. T. Fu, "Research on underwater object recognition based on YOLOv3," vol. 27, pp. 1837-1844, 2021.
- [68] J. Redmon and A. J. a. p. a. Farhadi, "Yolov3: An incremental improvement," 2018.
- [69] M. J. Islam *et al.*, "Semantic segmentation of underwater imagery: Dataset and benchmark," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 1769-1776: IEEE.
- [70] R. Parikh and N. J. A. a. S. Mehendale, "Detection of Underwater Objects in Images and Videos Using Deep Learning," 2023.
- [71] R. A. Dakhil and A. R. H. J. a. p. a. Khayeat, "Review On Deep Learning Technique For Underwater Object Detection," 2022.
- [72] F. Han, J. Yao, H. Zhu, and C. J. J. o. S. Wang, "Underwater image processing and object detection based on deep CNN method," vol. 2020, 2020.
- [73] L. Yihao, L. Qipei, Z. Yin, Z. Heyu, Z. Juntao, and C. J. 电. Xiang, "Review of underwater image object detection based on deep learning," vol. 45, no. 10, pp. 3468-3482, 2023.
- [74] S. Fayaz, S. A. Parah, G. J. M. T. Qureshi, and Applications, "Underwater object detection: architectures and algorithms—a comprehensive review," vol. 81, no. 15, pp. 20871-20916, 2022.
- [75] M. J. Islam, Y. Xia, J. J. I. R. Sattar, and A. Letters, "Fast underwater image enhancement for improved visual perception," vol. 5, no. 2, pp. 3227-3234, 2020.
- [76] M. J. Islam, P. Luo, and J. J. a. p. a. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," 2020.

## الخلاصة

يلعب كشف الكائنات تحت الماء دوراً حيوياً في تطبيقات متنوعة مثل استكشاف المحيطات ومراقبة البيئة والروبوتات تحت الماء. تقدم هذه الرسالة نهجاً مقترحاً لمعالجة تحديات كشف الكائنات تحت الماء، باستخدام مجموعة البيانات لتحليل الصور الفوتوغرافية الفوقية للبحريات (SUIM). يتم التركيز على تطوير نموذج مشفر-فك بالكامل مُضَبَّط بعناية، يحقق توازناً بين أداء الكشف وكفاءة الحوسبة. تسلط التحديات الفريدة لمراقبة واستكشاف المياه، بما في ذلك سوء الرؤية وتغيرات الظروف البيئية، الضوء على الحاجة إلى حلول متخصصة. يلعب تفريق المعاني دوراً حاسماً في هذا النهج، مع تعزيز قدرة النموذج على اكتشاف وتصنيف الكائنات بدقة تحت سطح الماء.

تستغل هندسة النموذج المقترح شبكة مشفر-فك بالكامل تستند إلى نموذج (VGG-16) لاستخراج معلومات فضائية معقدة من مشاهد تحت الماء. تستخدم الشبكة عملية تغيير حجم لمطابقة أحجام الإدخال والإخراج، مما يساعد في الحفاظ على الدقة الفضائية وتجنب فقدان المعلومات. تحتفظ هذه الشبكة بالسماوات العامة والمحلية، مما يساعد في اكتشاف الكائنات في ظروف تحت الماء متنوعة. ولتعزيز وضوح الكائنات المكتشفة بصرياً، يستخدم النموذج شبكة (ESRGAN) لتحسين وضوح الصور الفوتوغرافية تحت الماء ذات الدقة المنخفضة.

ومن أجل تحسين نتائج الكشف بشكل إضافي، يتم استخدام عمليات المورفولوجيا لإزالة الأجسام الفنية والضوضاء الصغيرة من التنبؤات، مما يؤدي إلى حدود أكثر دقة وتناسقاً بصرياً للكائنات. يُسهم دمج عمليات المورفولوجيا في خط الكشف في تحسين تحديد الموقع النهائي للكائنات.



جامعة كربلاء  
كلية علوم الحاسوب وتكنولوجيا المعلومات  
قسم علوم الحاسوب

## الكشف والتعرف على الكائنات تحت الماء باستخدام التعلم العميق

رسالة ماجستير  
مقدمة الى مجلس كلية علوم الحاسوب وتكنولوجيا المعلومات / جامعة كربلاء وهي جزء من  
متطلبات نيل درجة الماجستير في علوم الحاسوب

كتبت بواسطة  
رضوان عدنان داخل جبر

بإشراف  
أ.م.د. علي رضا حسون الخياط