International Journal of Mathematical, Engineering and Management Sciences
Vol. 9, No. 6, 1510-1530, 2024
https://doi.org/10.33889/IJMEMS.2024.9.6.080

Multiclass Image Segmentation using Deep Residual Encoder-Decoder Models in Highly Turbid Underwater Ambiances

T. P. Mithun Haridas
Center for Ocean Electronics Lab (CUCENTOL), Department of Electronics,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
Corresponding author: mithuntp@cusat.ac.in

Suraj Kamal
Center for Ocean Electronics Lab (CUCENTOL), Department of Electronics,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
E-mail: surajkamal@cusat.ac.in

Arun A. Balakrishnan
Center for Ocean Electronics Lab (CUCENTOL), Department of Electronics,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
E-mail: arunab@cusat.ac.in

Rosemol Thomas
Center for Ocean Electronics Lab (CUCENTOL), Department of Electronics,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
E-mail: rosemolthomas@cusat.ac.in

N. A. Nezla
Center for Ocean Electronics Lab (CUCENTOL), Department of Electronics,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
E-mail: nezla.latheef@cusat.ac.in

Kannan Balakrishnan
Department of Computer Applications,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
E-mail: bkannan@cusat.ac.in

M. H. Supriya
Center for Ocean Electronics Lab (CUCENTOL), Department of Electronics,
Cochin University of Science and Technology, 682021, Kochi, Kerala, India.
E-mail: supriya@cusat.ac.in

(Received on April 1, 2024; Revised on June 20, 2024 & August 15, 2024; Accepted on September 3, 2024)

Abstract

Underwater environments, especially the coral reefs, are the habitat of many critically endangered species. Extensive monitoring of these aquatic ecosystems is essential for conserving and deep understanding of these vulnerable habitats. Monitoring by extracting details from underwater images of turbid, hazy marine environments is extremely challenging. In this work, a novel annotated dataset is created for three classes of objects in the images of coral reef environment considering fish, rock/coral and background for the Fish4Knowledge dataset, a benchmark dataset primarily for binary segmentation. This work also proposes a multiclass ResUnet based image segmentation model for the newly created multiclass annotations. Various encoder-decoder convolutional architectures were analysed and found that ResUnet exhibits better robustness. The performance of the multiclass ResUnet model is also analysed by optimizing with different cost functions. Various underwater noisy conditions are simulated in the test images to find the robustness of the model, and observed that the proposed model optimised with Jaccard loss performs

better even in extremely noisy scenarios.

Keywords- Underwater image, Encoder-decoder model, Multi-class image segmentation, Fish4Knowledge, Underwater noise.

1. Introduction

Underwater imaging is an essential modality for various research fields including marine biology, environmental monitoring as well as exploration and navigation. The coral reef environment especially, possesses high biodiversity and is home to several vulnerable fishes, which can be indicator species for shifts in climatic conditions and pollution levels (Asner et al., 2020). Coral reef ecosystem in addition has high economic importance in terms of ecotourism (Yuan et al., 2024). Moreover, many of these fish species are on the verge of extinction as well, and the decline of such keystone species may even lead to ecosystem collapse (Asner et al., 2022; Mentzel et al., 2024). Close monitoring of these species and the environment can also throw light towards a broader understanding of such marine environments and ecosystems (Cael et al., 2023; Hoegh-Guldberg et al., 2017; Ortiz and Hermosillo-Núñez, 2024). However, manual efforts in continuous monitoring are often impractical and are prone to human errors. Hence, proper adoption of underwater imaging and computational techniques has to be utilised for close and careful analysis in these domains (Apprill et al., 2023; Horoszowski-Fridman et al., 2024).

In recent decades, there have been many advancements in underwater imaging techniques and data collection methods. However, the analysis of these images was found to be very challenging due to poor visibility, colour distortion, illumination variations, turbidity, haziness, noise and other artefacts that further distort the information. Light gets scattered in all directions when it interacts with the water molecules and suspended particles, which results in blurriness. They also cause reflection and refraction of light rays leading to the artefacts in brightness, which in turn results in the variation of image contrast. Absorption of longer wavelength of light by the water causes the suppression of red-region, disturbs the colour balance, and leads to spectral distortions (Duntley, 1963; Singh and Bhat, 2023; Zhou et al., 2023).

Manual analysis is impractical for monitoring very long video streams or footage from underwater exploration. Thus, image content analytic techniques such as detection, localisation, segmentation and tracking of various targets of interest can play a significant role in understanding underwater scenarios. Among them, image segmentation by virtue provides detection along with localisation, which can be extended to tracking. Image Segmentation involves partitioning an image into multiple segments or regions, and each segment represents a distinct object or region of interest within the image. Semantic Segmentation problem can be viewed fundamentally as a classification problem, where the learning model tries to predict the corresponding class of each pixel in an image. Hence, the multiclass semantic segmentation problem, which is the topic of interest in this work, boils down into a multiclass classification problem in the pixel space (Sakshi and Kukreja, 2023; Yu et al., 2023). By segmenting underwater images such as coral reefs, researchers and marine biologists can gain valuable insights into the marine ecosystem's health, biodiversity, and overall condition of the marine environment.

Several traditional image segmentation techniques including thresholding, clustering-based, edge-based, region-based and soft computing-based approaches have been utilized by the research community in various scenarios (Yadav and Pandey, 2022; Yu et al., 2023). Deep Learning techniques are found to outperform traditional techniques and proved to adapt better to different real-world scenarios with ambiguities, if sufficient amount of data is provided (Ghosh et al., 2019; Li et al., 2024; Minaee et al., 2022). Deep learning being a data-centric approach always demands a huge amount of labelled data so that models can learn all the possible distributions of data. Labelling and annotating data is a laborious and time-consuming task that needs careful attention (Bhagat and Choudhary, 2018). Data annotation for segmentation in the underwater scenario is drawing masks around the regions of interest, such as fish, coral reefs, and various marine species. The labelled data helps the model in learning to differentiate between different objects and backgrounds in the images.

In recent years, many studies were conducted for developing model for analysing coral reef environment with the help of imaging techniques (Li et al., 2024), especially with deep learning approach. Remote sensing based study with satellite and ariel imaging was used widely for coral reef monitoring and surveys.

(Giles et al., 2023; Lyons et al., 2024). Habitat mapping and classification of corals from the hyper-spectral images was discussed by Rashid and Chennu (2020) and semantic mapping for analysing coral reef variations in climate change was proposed by Zhong et al. (2023). Most of the existing researches lacks close monitoring of coral ecosystem with special focus on the fish population, which is the key to indicate the health and condition of the ecosystem. Moreover, the impact of underwater noise in the images is also crucial in the developing a dependable system, which is also lacking in the literature.

This paper aims to analyse ResUnet-based multiclass semantic image segmentation on underwater images with multiclass scenarios in the coral reef environment with fish, rock, and background as different classes. The Fish4knowledge dataset is adapted for multiclass semantic segmentation by creating annotations for three classes of pixels. Multiclass semantic segmentation was carried out using U-Net architecture with the encoder backbone utilizing residual modules. The model is quantitatively analysed using evaluation metrics like Intersection over Union (IoU), Accuracy, Precision, and Recall. Performance on various simulated underwater noise scenarios is also observed to understand the robustness of the model. Thus, this research contributes to fill the existing research gaps by developing deep learning model for close monitoring of the coral reefs with multiclass objects including rock and fish. Moreover, the work also contributes through the analysis of performance of the models with regard to the deterioration that might happen due to the impact of prominent underwater noises.

2. Related Works

Underwater Image analysis has been researched elaborately for the past several years. Some of the most noted research works related to underwater datasets and deep learning-based approaches for the analysis of underwater datasets are discussed here.

2.1 Underwater Image Dataset

A few underwater fish datasets are available in the literature. Some of the existing benchmark dataset details are as given in the **Table 1**. Fish4Knowledge Underwater datasets consist of video footage of fish assemblage in coral reefs and three ground truth datasets for problems such as target detection, fish species recognition and trajectory-based fish behaviour analysis (Boom et al., 2012a; Boom et al., 2012b). Rock Fish dataset consist of images of rockfish near seabed collected by ROV for rockfish species recognition (Stierhoff and Cutter, 2013). QUT fish dataset (Anantharajah et al., 2014) consists of 3960 images from in situ and controlled environment for classification task. Deep Fish dataset consists of images from complex underwater fish habitats for problems such as fish count, identifying their locations, and estimate their sizes (Saleh et al., 2020). SUIM dataset consists of images collected during ocean exploration containing annotations of eight different object categories (Islam et al., 2020).

Table 1. Underwater fish dataset characteristics.

Dataset name	No. of images	$Task	Resolution	Environment/ Collection method	Segmentation classes
Rockfish (Stierhoff and Cutter, 2013)	4307	D	1280×720	Insitu	—
QUT (Anantharajah et al., 2014)	3960	CL	480×360	Insitu and Controlled	—
Deep Fish (Saleh et al., 2020)	39766	CL, CN, L, S	1920×1080	Insitu	2
SUIM (Islam et al., 2020)	1500	S	#varying sizes	Ocean exploration	8$^{\alpha}$
Fish4Knowledge (Boom et al., 2012a)	27370	CL, L, S	~128×128	Insitu	2
*Ours	1200	S	128×128	Insitu	3*

***Ours**: *Created multi-class annotation for the selected images of Plectroglyphido don dicky class in the Fish4Knowledge database with three class annotations including Fish, Rock and Background pixel classes*
$Tasks: *CL - Classification, CN - Counting, D - Detection, L - Localisation, S - Segmentation*
#varying sizes: *1906 × 1080, 1280 × 720, 640 × 480, and 256 × 256*
~128×128: *Images have varying sizes in the database, but resized to 128 × 128 for processing*
8$^{\alpha}$: *Eight classes including fish, reefs, aquatic plants, wrecks/ruins, human divers, robots, sea-floor, and background*

2.2 Deep Learning Based Analysis

Encoder-Decoder architectures of CNN were found to have capabilities for better image segmentation and classification (Ji et al., 2021). These approaches outperform conventional techniques and traditional image processing algorithms. Over the years, extensive studies on semantic segmentation-based approaches haven been conducted in many computer vision problems such as Crowd Tracking and Anomaly Detection (Abdullah and Jalal, 2023), Leaf Diseases Detection (He et al., 2024), Retinal Vessel Segmentation (Yakut et al., 2023), Fully Convolutional Neural Networks for Improved Brain Segmentation (Khaled et al., 2023). Several studies on the underwater ecosystem were also carried out utilizing deep learning architectures for fish detection and recognition (Cui et al., 2020; Li et al., 2015; Moniruzzaman et al., 2017). Analysis of coral reef fishes by a supervised machine learning model was proposed (Villon et al., 2016) and observed better performance over Support Vector Machine (SVM) and Histogram of Oriented Gradients (HOG) methods. Islam et al. (2020) proposes a deep residual model with skip blocks for segmenting underwater images from ocean exploration with eight classes. Mizuno et al. (2020) has proposed a coral survey method using U-net architecture with the help of optical camera system, and Song et al. (2021) proposes a classification model DeeperLabC, for coral and non-coral areas with single channel data. Thomas et al. (2022) combines two distinct U-Net models to handle RGB images and grayscale images separately and reported better feature extraction. Classification methods of seabed images in coral reefs was also developed in recent years (Jackett et al., 2023) using the ResNet50 architecture.

U-Net based convolutional encoder-decoder architectures were found to be effective for segmenting images, particularly those from the medical domain (Ronneberger et al., 2015). The encoder extracts the context, while the decoder consists of the up-sampling path that produces the accurate localized segmentation map. U-Net based supervised segmentation (Nezla et al., 2021) and analysis of U-Net based segmentation on five different fish species (Thampi et al., 2021) on underwater fish images from Fish4Knowledge dataset were also conducted in recent years.

3. Dataset Creation

3.1 Multiclass Data Annotation

A novel multiclass dataset annotation is created utilizing the selected images of Plectroglyphido don dicky class in the Fish4Knowledge dataset (Boom et al., 2012a; Boom et al., 2012b) with coral reef environment. This study was conducted, as a model system for the analysis of coral reef environment and, as a part of a broader study regarding habitat loss and keystone species analysis. In this regard, Fish4Knowledge dataset was found to be more relevant for the study and hence selected as the candidate dataset for this work.

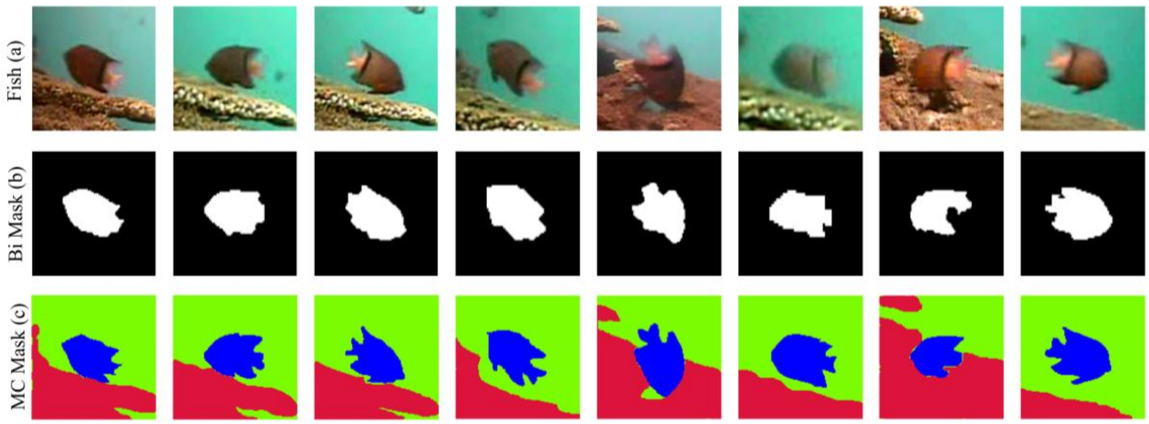


Figure 1. Sample dataset showing: (a) fish images, (b) original binary annotations Bi Mask from Fish4Knowledge dataset (Boom et al., 2012a), and (c) proposed multiclass annotations MC Mask.

Fish4Knowledge Dataset is an in-situ dataset with low-quality images with 23 classes of fish species, primarily for binary segmentation of fish from underwater coral reef scenarios. This research extends the dataset by creating multi-class annotations, with classes fish, rock, and background. Three-class annotated mask samples, along with their input images and original binary masks, are shown in **Figure 1**. Data preparation for multiclass semantic segmentation is carried out using the Apeer annotation tool (Apeer, 2023).

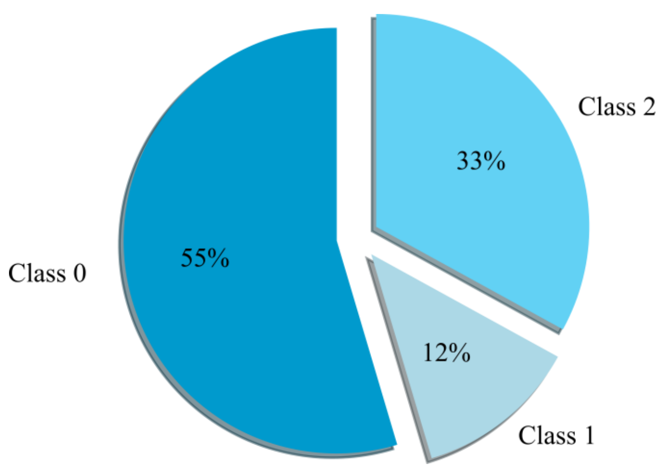


Figure 2. Data distribution of pixel-level data class.

Apeer is an intuitive annotation tool for deep learning needs those aids in annotating multidimensional datasets. Pixels in each image are classified into three categories: Fish, Rock, and Background class. The semantic segmentation annotation option is utilized for creating the annotations. Background class is predefined, and two more classes of fish and rock are annotated and masks are exported as tiff images. The class distribution of background, fish, and rock is very unbalanced, and hence, a weighting matrix is used for processing the images. Class distribution, hex-code for class annotations, and encoded class values are as shown in **Table 2**. The data distribution indicating the pixel-level data classes is shown in **Figure 2**.

Table 2. Proposed data annotation details of three classes.

Class	Background	Fish	Rock
Class Distribution[1]	10730739	2441037	6489024
Annotation Color[2]	#6900ff	#5f7ee2	#df7a5e
Encoded Value[3]	0	1	2

[1]*Distribution of pixel classes for the entire dataset (pixel count)*
[2]*Hex color code value used for pixel class annotation*
[3]*Label encoded value for the pixel classes*

3.2 Data Pre-processing
The Fish4Knowledge data set contains images of varying sizes. The selected images of Plectroglyphido don dicky class images are standardized to the size of 128×128. Input image data is reshaped to 128×128×3 and scaled from [0:255] to [0:1]. The mask is reshaped and converted to categorical values by applying label encoding to make pixel labels as 0, 1, 2 for the background, fish, and rock classes respectively. The dataset created consists of 1200 image annotations. The training set comprises 70% of the total data while remaining equally divided for validation (15%) and testing (15%).

4. Methodology
4.1 Multiclass ResUnet Model
The network architecture for multiclass semantic segmentation is adapted from the U-Net architecture (Ronneberger et al., 2015). The encoder section of the model consists of a series of convolutional, batch normalisation and maxpool layers. The decoder section consists of up-sampling and transpose convolutions. The residual block provides skip connections between a series of operations, thereby incorporating identity mapping and reducing vanishing gradients (He et al., 2016).

Table 3. Network structure of the multiclass ResUnet architecture for the segmentation of three-class Fish4Knowledge annotation data.

Section	Residual units	Conv layer	Filter size	Channel	Stride	Output size
Input						128x128x3
Encoding	Unit 1	Conv 1	3x3	64	1	128x128x64
		Conv 2	3x3	64	1	128x128x64
	Unit 2	Conv 3	3x3	128	2	64x64x128
		Conv 4	3x3	128	1	64x64x128
	Unit 3	Conv 5	3x3	256	2	32x32x256
		Conv 6	3x3	256	1	32x32x256
Bridge	Unit 4	Conv 7	3x3	512	2	16x16x512
		Conv 8	3x3	512	1	16x16x512
Decoding	Unit 5	Conv 9	3x3	256	1	32x32x256
		Conv 10	3x3	256	1	32x32x256
	Unit 6	Conv 11	3x3	128	1	64x64x128
		Conv 12	3x3	128	1	64x64x128
	Unit 7	Conv 13	3x3	64	1	128x128x64
		Conv 14	3x3	64	1	128x128x64
Output		Conv 15	1x1	3	1	128x128x3

ResUnet utilizes residual blocks in the encoder section of U-Net-based encoder-decoder architecture to get better feature extraction (Zhang et al., 2018). ResUnet takes the input to a sequence of convolutional (Conv), batch normalisation (BN), ReLU, Conv and then add with input using the residual skip connection. Thereafter a sequence of residual blocks with a set of BN, ReLU, Conv, BN, ReLU, Conv followed by addition to the input to the residual block are utilised for the encoder and decoder. In the decoder section, the encoder features are first up-sampled and then concatenated with the output of the corresponding encoder section. ResUnet architecture, designed for binary segmentation, is adapted for multiclass scenarios by incorporating a three channel Conv layer followed by a SoftMax layer in the decoder section. The architecture of multiclass ResUnet model is shown in **Figure 3**. The Conv filter size and activation function details of the multiclass ResUnet architecture is shown in **Table 3**.

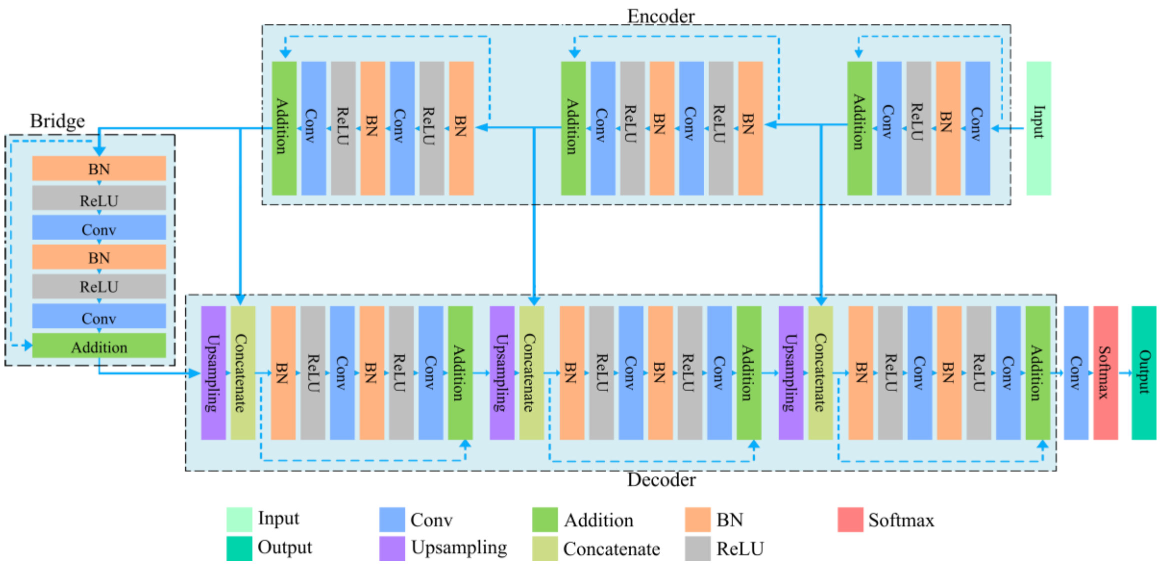


Figure 3. The architecture of the multiclass ResUnet model.

4.2 Different Encoder Backbone Architectures

Different CNN architectures are analysed by changing the encoder section of U-Net segmentation architectures such as basic U-Net encoder, VGGnet (VGG19), Inception v3 and ResUnet architecture. The multiclass segmentation models are compared with the segmentation performance using the binary segmentation images of the Fish4Knowledge dataset using U-Net. The capability of segmentation architectures in identifying the fish characteristics from the image are compared and it has been observed that the addition of the new class improved the segmentation of the fish to a better extend.

Table 4. Performance comparison of different models.

Model	Dataset type	Class	IoU	Mean IoU
U-Net	Binary segmentation	Class 0	0.9768	0.9159
		Class 1	0.8550	
U-Net	Multiclass segmentation	Class 0	0.9285	0.8928
		Class 1	0.8398	
		Class 2	0.9102	
U-Net with VGG19 encoder	Multiclass segmentation	Class 0	0.9340	0.8996
		Class 1	0.8475	
		Class 2	0.9172	
U-Net with inception v3 encoder	Multiclass segmentation	Class 0	0.9406	0.9090
		Class 1	0.8619	
		Class 2	0.9244	
ResUnet ($\mathcal{L}_{CL}$)	Multiclass segmentation	Class 0	0.9523	**0.9221**
		Class 1	**0.8755**	
		Class 2	0.9386	

4.3 Loss Functions

Background class pixels are higher than the other two classes, as described in the **Table 2**. Out of the total data, 55% is of class 0 (background), 12% of class 1 (fish) and 33% of class 2 (rock). Four different loss functions are used separately to train the model and the performance of the multiclass ResUnet model is analysed with each loss function.

***Cross-Entropy Loss*: ** Multiclass classification is achieved by incorporating the softmax layer in the output layer as given in the equations below. Considering C number of classes, the probability of the i^{th} class is as given the Equation (1). Cross-Entropy (CE) measures the dis-similarity between the predicted and the ground truth classes. CE loss equation is as given in Equation (2) where d_i is the ground truth or desired response and p_i is the predicted response.

$$f(p)_i = \frac{e^{p_i}}{\sum_j^C e^{p_j}} \tag{1}$$

$$loss_{CE} = -\sum_{i=1}^{C} d_i \log(f(p)_i) \tag{2}$$

***Focal Loss*:** Focal loss incorporates weights to the contribution of each sample to the loss based on the classification error. Pixel classes that are correctly classified will be weighted less, thus decreasing their contribution to the loss. Focal loss is represented using Equation (3), where α is the weighting factor with a value of 0.25 and γ is the focusing parameter with a value of 2.0.

$$loss_{FL} = -\sum_{i=1}^{C} \alpha(1-p_i)^{\gamma}\, d_i \log(p_i) \tag{3}$$

***Dice Loss*:** Dice loss is the measure of similarity between predicted and ground truth segmentation masks. Dice loss incorporates the coefficient of precision (P), recall (R) and balance (β) with a value of 1.0 as in Equation (4).

$$loss_{Dice} = 1 - (1+\beta^2)\frac{P.R}{\beta^2.(P+R)} \tag{4}$$

***Jaccard Loss*:** Jaccard loss measures the similarity with the help of the Jaccard Index. i.e., the Intersection over Union (IoU) between prediction and ground truth mask, as in Equation (5), using precision (P) and recall (R).

$$loss_{Jaccard} = 1 - \frac{(P \cap R)}{(P \cup R)} \tag{5}$$

4.4 Optimiser

Adam is used as an optimizer with an initial learning rate of 1×10^{-3} (Kingma and Ba, 2014). The learning rate will be reduced 10^{-1} times if the validation loss is not reducing for five successive epochs. An early stopping criterion is set if the validation loss is not reduced for 20 successive epochs. The minimum learning rate was set as 1×10^{-7}.

4.5 Evaluation Metrics

The model performance is evaluated using loss, accuracy, and IoU score in the training and validation phase. The performance of the model during testing is assessed with the test data using metrics such as Precision, Recall, F1 score, and IoU scores.

Table 5. Quantitative analysis of training and validation on the multiclass ResUnet model with different loss functions.

Model	Architecture	Loss function	Metric	Training	Validation
$\mathcal{L}_{CL}$	ResUnet	Categorical Cross Entropy Loss (CL)	Accuracy	0.9643	0.9578
			IoU	0.8803	0.8692
			Loss	0.0296	0.0364
$\mathcal{L}_{FL}$	ResUnet	Categorical Focal Loss (FL)	Accuracy	0.9619	0.9576
			IoU	0.7566	0.7536
			Loss	0.0024	0.0029
$\mathcal{L}_{DL}$	ResUnet	Dice Loss (DL)	Accuracy	0.9609	0.9578
			IoU	0.9084	0.9009
			Loss	0.0483	0.0521
$\mathcal{L}_{JL}$	ResUnet	Jaccard Loss (JL)	Accuracy	0.9645	0.9578
			IoU	0.9162	0.9014
			Loss	0.0836	0.0977

5. Model Training and Analysis

U-Net model (Ronneberger et al., 2015) was trained on the original binary segmentation Fish4Knowledge dataset and later, the proposed three-class annotated segmentation dataset was utilised to train the U-Net model and the U-Net model with different encoder backbones with VGG19 (Simonyan and Zisserman, 2014) and Inception-v3 (Szegedy et al., 2016). These results were compared with the results of the multiclass ResUnet model($\mathcal{L}_{CL}$) which was also trained on the proposed three-class segmentation dataset. All the models were trained utilising the Cross-Entropy loss function. It was observed that the class 1 (fish) was recognised better by multiclass ResUnet architecture when compared to all other models. Considering the binary dataset, the addition of the rock class in the newly annotated three-class segmentation dataset improved the fish class's segmentation performance to a better extent. The model with Inception v3 backbone had a similar but slightly lower performance with ResUnet. The better performance exhibited by ResUnet may be due to the advantage of the residual connections in its architecture. Comparative results of different models in the test data are given in the **Table 4**.

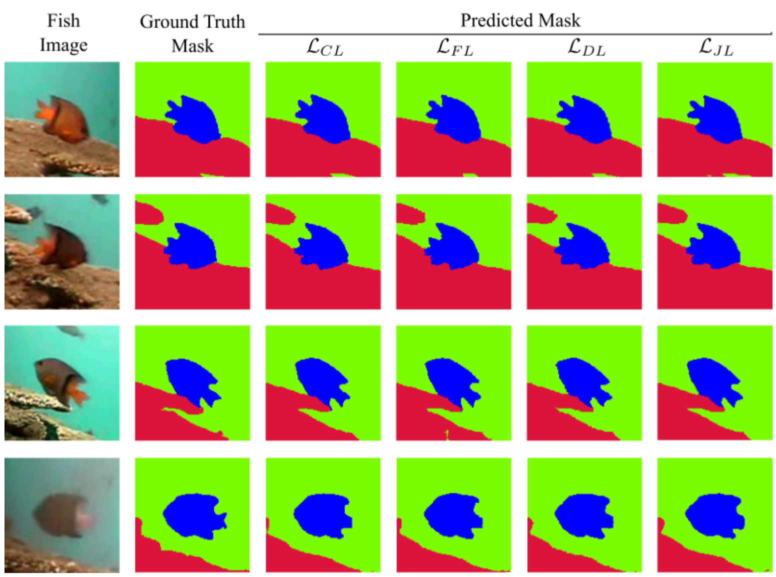


Figure 4. Segmentation mask predicted by multiclass ResUnet model optimised with four different loss functions Categorical Cross-Entropy loss ($\mathcal{L}_{CL}$), Categorical Focal loss ($\mathcal{L}_{FL}$), Dice loss ($\mathcal{L}_{DL}$) and Jaccard loss ($\mathcal{L}_{JL}$).

For further analysis, ResUnet architecture is trained on the proposed three-class annotated dataset with four different loss functions. ResUnet Model optimized with Categorical Cross-Entropy loss function is denoted as $\mathcal{L}_{CL}$, with Categorical Focal loss as $\mathcal{L}_{FL}$, Dice Loss as $\mathcal{L}_{DL}$, and Jaccard loss function as $\mathcal{L}_{JL}$ respectively. The number of trainable parameters of the multiclass ResUnet model is 8,221,123 and non-trainable parameters is 6,400. Optimisation pipeline is implemented using tensorflow libraries. Model is trained on NVIDIA GeForce RTX 3060 and requires 24ms for the end to end processing on inference per image, corresponding to a rate of 41.66 frames-per-second (FPS), which will be suitable for real-time underwater applications. The training performance of each model is analysed with accuracy, loss, and IoU scores. It is observed that $\mathcal{L}_{JL}$ obtained the highest accuracy and IoU scores, whereas the lowest accuracy and IoU scores were obtained by $\mathcal{L}_{DL}$ and $\mathcal{L}_{FL}$ respectively as shown in **Table 5**. It was analysed that the $\mathcal{L}_{FL}$ converges faster followed by the ResUnet model optimised with $\mathcal{L}_{CL}$, $\mathcal{L}_{DL}$, and $\mathcal{L}_{JL}$. Except for Focal loss all other losses have large variations in the loss values with the validation data.

6. Results and Discussions

6.1 Segmentation Results on Test Data

The optimized models are analysed for their performance with the test data. The segmented mask produced by four models is shown in **Figure 4**. The proposed multiclass ResUnet model with all four loss functions are found to have comparable performance. However, $\mathcal{L}_{JL}$ with Jaccard loss function shows the upper hand in producing accurate segmentation masks. For careful analysis of models, the confusion matrix of the best and worst performing images for each model is also analysed.

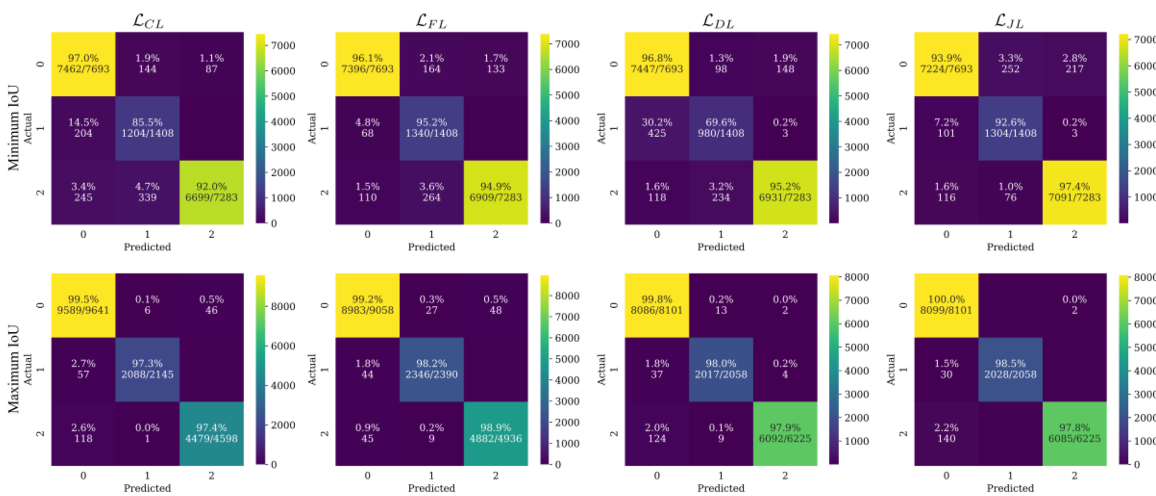


Figure 5. Analysis of confusion matrix of images with maximum and minimum IoU for the multiclass ResUnet model with different loss functions; $\mathcal{L}_{CL}$, $\mathcal{L}_{FL}$, $\mathcal{L}_{DL}$ and $\mathcal{L}_{JL}$.

6.2 Segmentation Performance of Different Models on a Single Image

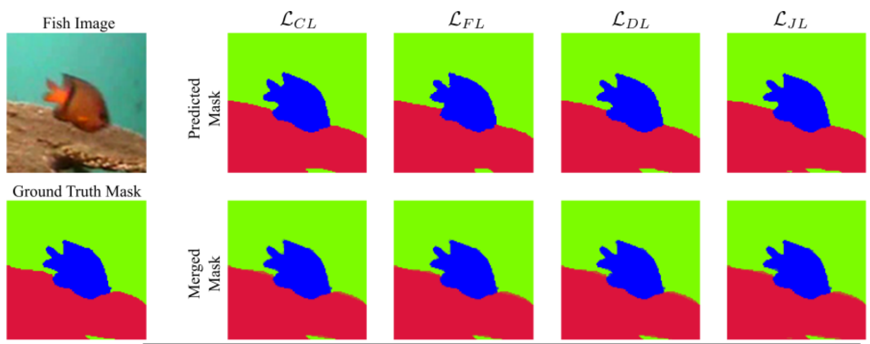
It was observed that each model performance varies with different images and the images that give best and worst IoU score also varies in different models. Hence, in order to compare the segmentation performance among the models one of the least and best performing common images is selected for analysis.

The segmentation mask produced by all four multiclass ResUnet models are analysed with one of the images with the highest and lowest IoU scores. The segmented mask produced is compared with the ground truth and the intersection of the ground truth mask with the predicted mask is also plotted as shown in

Figure 6 and **Figure 7**. It is observed that on the images with good and moderate resolution, all model performs well, with $\mathcal{L}_{CL}$ and $\mathcal{L}_{JL}$ having better performance. For low-resolution images, models $\mathcal{L}_{CL}$ and $\mathcal{L}_{FL}$ cannot perform well, whereas $\mathcal{L}_{FL}$ and $\mathcal{L}_{JL}$ perform moderately well, with $\mathcal{L}_{JL}$ having better performance. Thus, Focal loss and Jaccard loss performed well for ambiguous images with interclass interference in the pixel features, and all four losses Categorical Cross-Entropy, Categorical Focal loss, Dice loss, and Jaccard loss perform well in moderate ambiguous images. However, the Jaccard loss was found to be performing better with all the images across all the cases.

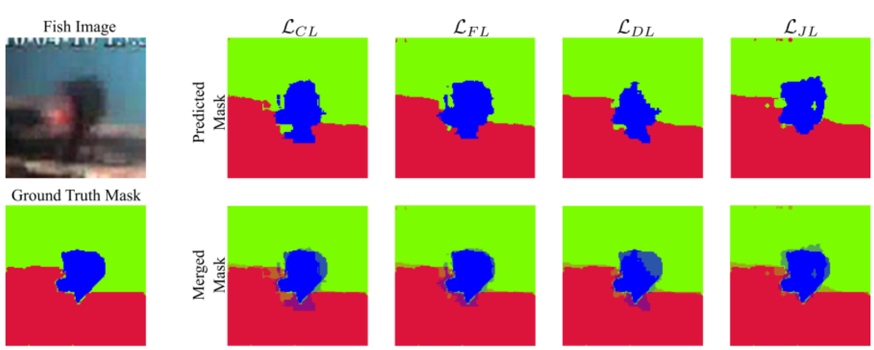
6.3 Confusion Matrix Analysis on Interclass Ambiguities

Confusion Matrix of images, which resulted in the maximum and minimum IoU for all four models, are analysed in **Figure 5**. It is observed that the test image with the best IoU score shows more ambiguity between classes fish (1) and background (0) than fish (1) and rock (2) for all the models. Jaccard loss-based $\mathcal{L}_{JL}$ shows the best results among all. When analysed with the images having the least IoU score, it is observed that the pixel level classification shows ambiguities in the classes 0 and 1 for the loss function $\mathcal{L}_{CL}$, $\mathcal{L}_{FL}$ and $\mathcal{L}_{DL}$, whereas $\mathcal{L}_{JL}$ observes ambiguities among all three classes. However, the deviations in least IoU images are outlier cases, which are analysed in detail using the IoU boxplot in **Figure 8**.



Mean IoU	0.9869	0.9869	0.9884	0.9895
Class0 IoU	0.9762	0.9775	0.9786	0.9792
Class1 IoU	0.9603	0.9568	0.9697	0.9854
Class2 IoU	0.9762	0.9757	0.9776	0.9771

Figure 6. Segmentation mask analysis on the test image that observed the highest IoU score.



Mean IoU	0.9378	0.9548	0.9373	0.9533
Class0 IoU	0.9164	0.9396	0.9042	0.9132
Class1 IoU	0.6367	0.7298	0.5632	0.7511
Class2 IoU	0.9089	0.9316	0.9323	0.9450

Figure 7. Segmentation mask analysis on the test image that observed the lowest IoU score.

6.4 Analysis of Evaluation Metrics

The box plot shown in **Figure 8** represents the mean IoU and class-wise IoU scores of class 0 (background), class 1 (fish) and class 2 (rock) classes. It is observed that $\mathcal{L}_{JL}$ optimised with the help of the Jaccard loss function performs better than all other models. However, $\mathcal{L}_{DL}$ optimized with Dice loss function also shows a similar IoU score, but the performance is inconsistent as $\mathcal{L}_{DL}$ possesses outliers in the prediction. Considering the class fish, the total number of outliers observed are 4 and 5 for $\mathcal{L}_{JL}$ and $\mathcal{L}_{DL}$ respectively. $\mathcal{L}_{CL}$ optimized with Categorical Cross-Entropy loss function and $\mathcal{L}_{FL}$ optimised with Categorical Focal loss function shows low performance compared to $\mathcal{L}_{DL}$ and $\mathcal{L}_{JL}$. A total number of 6 outliers are observed for $\mathcal{L}_{CL}$ and $\mathcal{L}_{FL}$ each. It is also noticed that the worst performance of all four model loses are on the same fish image as analysed in the **Figure 7**. It is observed that even the blurred image has been segmented with a class 1 IoU of 0.751 by $\mathcal{L}_{JL,}$ which has the highest performance, followed by $\mathcal{L}_{FL}$, $\mathcal{L}_{CL}$ and $\mathcal{L}_{DL}$ with a class 1 IoU of 0.729, 0.636 and 0.563 respectively.

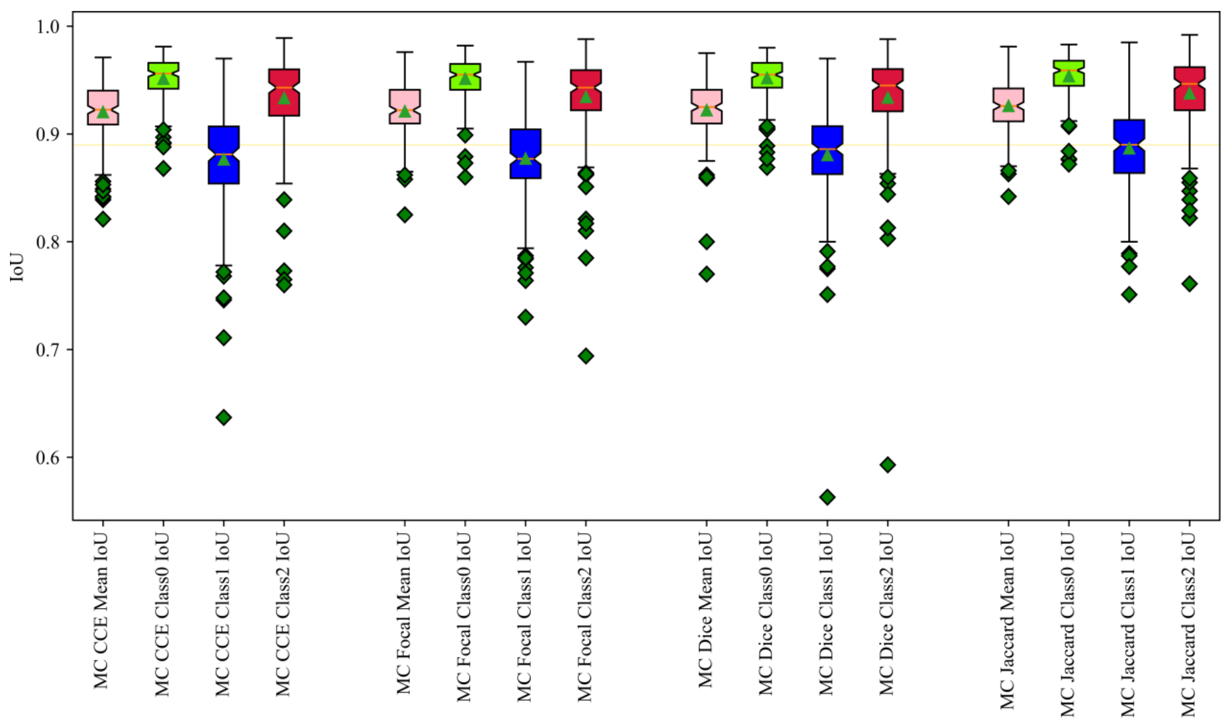


Figure 8. Analysis of Multi-Class (MC) segmentation performance comparing the IoU of the multiclass ResUnet model with different loss functions such as Categorical Cross-Entropy loss (CCE), Categorical Focal loss, Dice loss, and Jaccard loss.

Quantitative evaluation of the overall performance of the four models is analyzed with the help of the metrics precision, Recall, F1 Score, class-wise IoU, and mean IoU score, considering all the classes together in the test data as shown in the **Table 6**. Considering the main target of interest (fish), $\mathcal{L}_{JL}$ and $\mathcal{L}_{DL}$ are found to have the best class 1 precision of 0.9434 and $\mathcal{L}_{CL}$ is found to have the best class 1 recall of 0.9365. $\mathcal{L}_{JL}$ has the best F1 score of 0.9396, class 1 IoU of 0.8861, and mean IoU score of 0.9279. In conclusion, it can be observed that $\mathcal{L}_{JL}$ is found to outperform all other models in terms of better precision, F1 score, class 1 IoU, and mean IoU score considering all the classes.

The proposed multiclass ResUnet model shows comparable segmentation results with the state-of-the-art models in similar domains having multiclass underwater image segmentation dataset. UISS-Net (He et al., 2024) evaluated on SUIM dataset achieved a mIoU of 72.09, FSFS-Net (Yang et al., 2023) achieved a mIoU of 79.62 on fish feeding datasets, CNet model (Zhang et al., 2024) on segmentation of coral seabed images achieved mIoU of 81.83, and U-Net-s Rns269e model (Chicchon et al., 2023) evaluated on dataset with fish, sea floor and water classes created out of images from SUIM, Rock fish and Deep fish datasets achieved mIoU of 87.45. Our proposed model ($\mathcal{L}_{JL}$) also shows better results with three-class annotation for coral reef images from Fish4Knowledge dataset with a mIoU value of 92.79.

Table 6. Quantitative analysis of pixel classification report on test data.

Model	Metric	Support	Precision	Recall	F1 Score	IoU	Mean IoU
$\mathcal{L}_{CL}$	Class 0	1632837	0.9714	0.9797	0.9756	0.9523	0.9221
	Class 1	379504	0.9308	**0.9365**	0.9336	0.8755	
	Class 2	936779	0.9769	0.9599	0.9683	0.9386	
$\mathcal{L}_{FL}$	Class 0	1632837	0.9731	0.9778	0.9754	0.9520	0.9229
	Class 1	379504	0.9390	0.9294	0.9342	0.8765	
	Class 2	936779	0.9713	0.9670	0.9691	0.9401	
$\mathcal{L}_{DL}$	Class 0	1632837	0.9730	0.9784	0.9757	0.9526	0.9242
	Class 1	379504	**0.9434**	0.9291	0.9362	0.8801	
	Class 2	936779	0.9708	0.9674	0.9691	0.9400	
$\mathcal{L}_{JL}$	Class 0	1632837	0.9748	0.9788	0.9768	0.9546	**0.9279**
	Class 1	379504	**0.9434**	0.9358	**0.9396**	**0.8861**	
	Class 2	936779	0.9727	0.9688	0.9707	0.9431	

6.5 Analysis of Performance over Various Underwater Noisy Conditions

Performance of the model with respect to the different probable noisy conditions in the underwater scenarios such as blurriness, brightness reduction and red channel reduction of the images are analysed using synthesised images (Duntley, 1963; Galdran et al., 2015; Peng and Cosman, 2017). An image in the test data is analysed for three underwater noise cases at various levels. The blurriness of the image is analysed by applying the gaussian blur function with various window sizes corresponding to the lowering of the Structural Similarity Index Measure (SSIM). SSIM is utilised for analysing the quality of the noisy image with respect to the original image in the dataset (Brunet et al., 2012; Wang et al., 2004). Similarly, noisy scenarios in the brightness and red channels are simulated by the reduction of brightness and red channel at various levels, which are more prominent scenarios in underwater imaging.

Figure 9 shows the noisy images, histogram of the corresponding noisy images, and the predicted segmentation mask in the aforementioned scenarios for the multiclass ResUnet model optimised with different loss functions ($\mathcal{L}_{CL}$, $\mathcal{L}_{FL}$, $\mathcal{L}_{DL}$ and $\mathcal{L}_{JL}$). Similarly, **Figure 10** shows the analysis of variations of IoU using different loss functions and various noisy scenarios. It is observed that increasing blurriness in the images primarily affects class 1 (fish) and has less influence on other classes. The model optimised with Jaccard loss is found to be more robust against the blurred conditions and maintains its IoU performance of 0.8 untill the blurriness level of window size 45 i.e. SSIM 0.5.

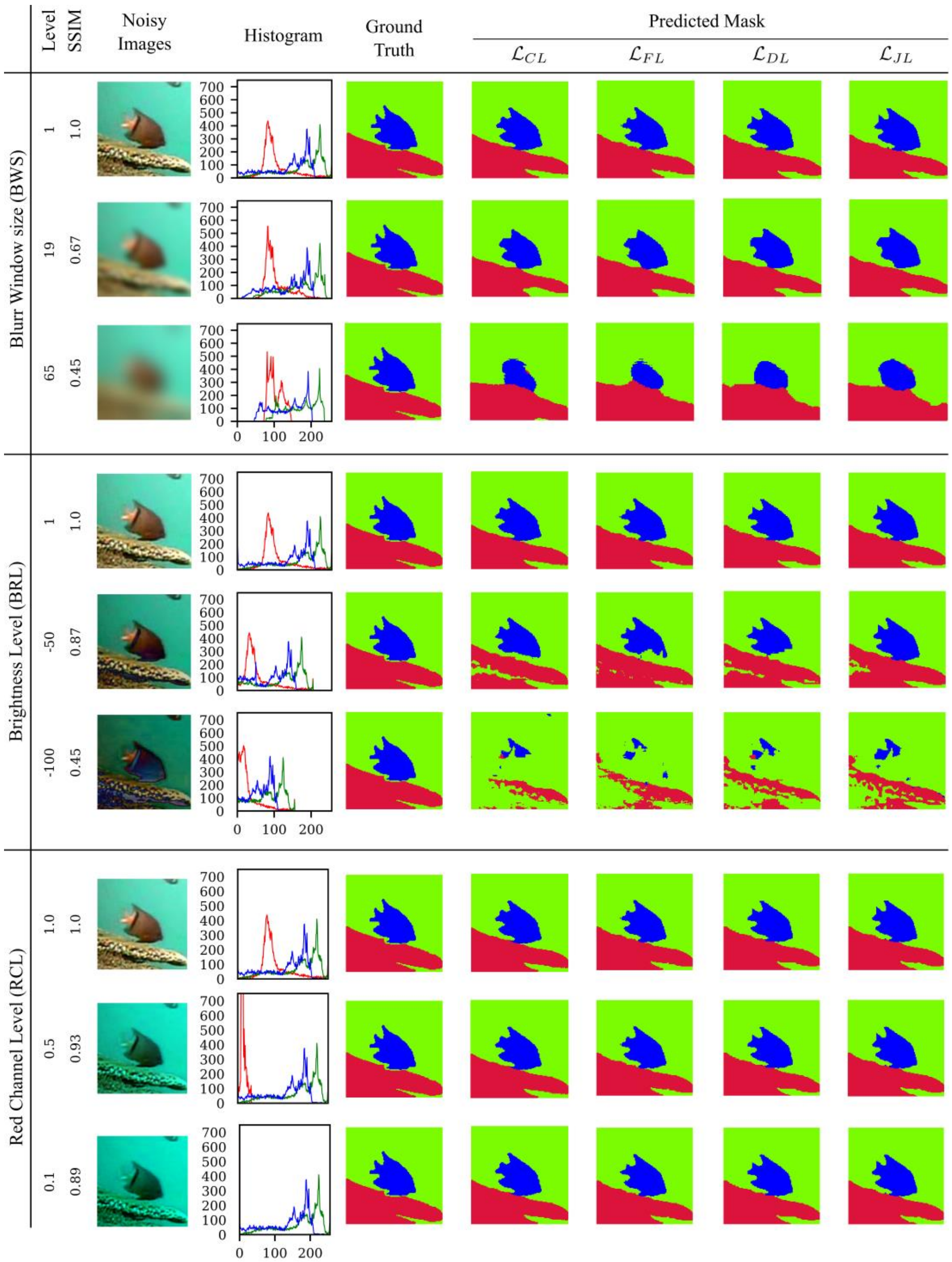


Figure 9. Analysis of model performance with different underwater noisy scenarios. Noisy conditions considered are: (1) Gaussian Blurr effect with various blur window sizes (BWS), (2) Brightness Reduction with various brightness Levels (BRL), and (3) Red Channel Reduction with various Red Channel Levels (RCL). The image quality at each noisy level is indicated with the metric SSIM. The predicted mask of multiclass ResUnet model optimised with four different loss functions $\mathcal{L}_{CL}$, $\mathcal{L}_{FL}$, $\mathcal{L}_{DL}$ and $\mathcal{L}_{JL}$ are also analysed and observed that Jaccard loss ($\mathcal{L}_{JL}$) exhibits robust performance.

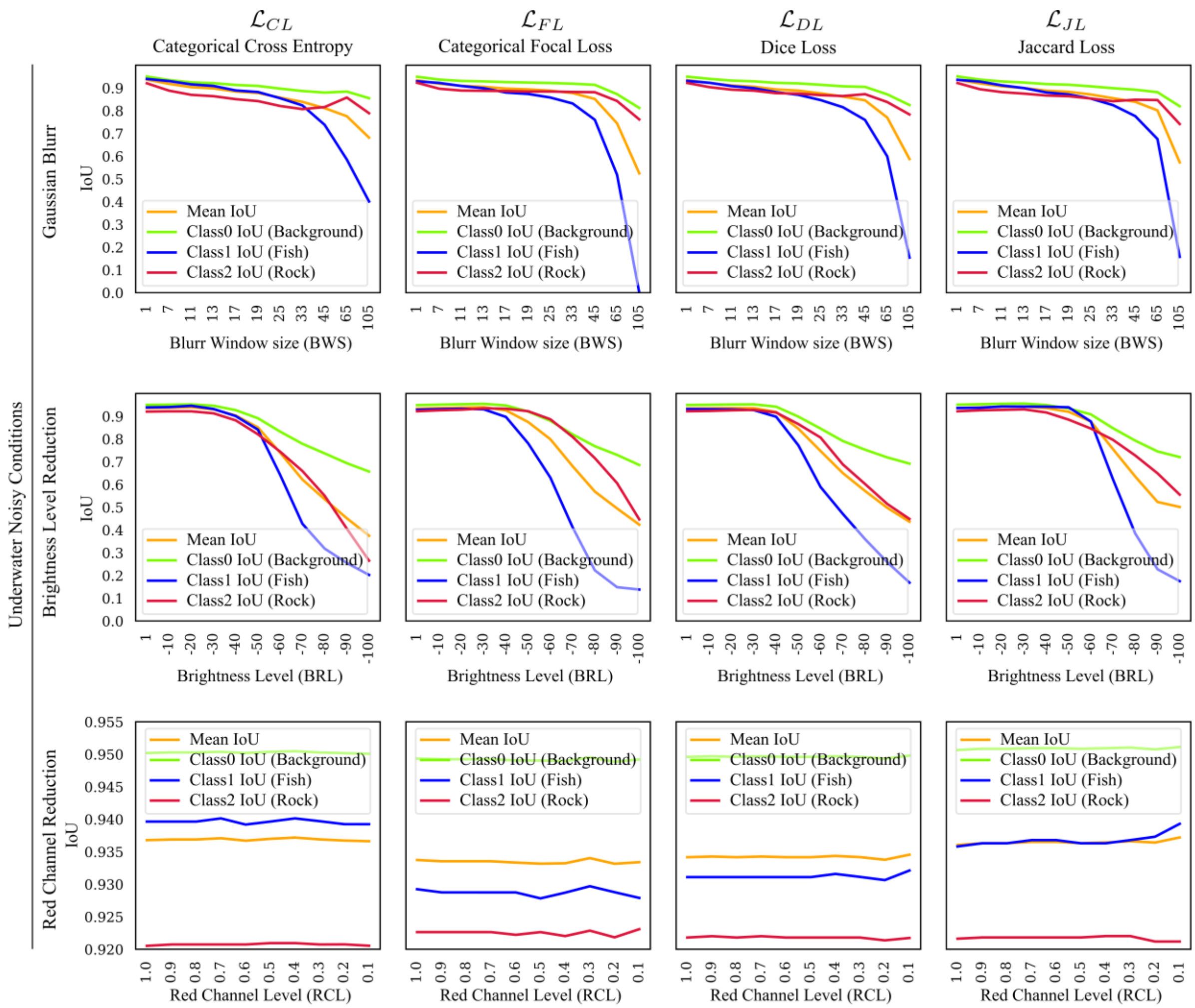


Figure 10. Analysis of the performance of each loss function on different noisy scenarios with respect to IoU.

Blurriness causes the distribution of pixels of fish more towards either of the classes, rock or background, causing the decrease in the IoU score of fish, as the foreground objects tend to blend with the background. In the model optimised with Jaccard loss, the recall values of the fish decrease below 0.8 after the blurriness level of window size 45, i.e. SSIM 0.5, with precision and recall corresponding to the blurriest image being 1 and 0.1545. At the same time, the class 2 rock maintains precision in the range of 0.99 to 0.78 and recall in the range of 0.92 to 0.85. Similarly, background maintains precision in the range of 0.96 to 0.86 and recall in the range of 0.9 to 0.93. From the observation, it can be understood that fish, the primary target of interest, possesses well-defined structures such as edges compared to the rock and background. Hence, as the blurriness increases the fish pixels get more confused with rock and background.

Reduction in the brightness levels affects the performance of all three classes, with fish being the most affected. From the IoU plot it can be observed that $\mathcal{L}_{JL}$ (Jaccard loss) was found to be robust compared to all other loss functions i.e. IoU score of 0.9 till the brightness level of -60 with SSIM 0.81. Whereas $\mathcal{L}_{CL}$, $\mathcal{L}_{FL}$ and $\mathcal{L}_{DL}$ were able to hold their performance at the brightness levels -40, -30 and -30 respectively. In the Jaccard loss optimised model, the recall values of the fish decrease below 0.9 after brightness level -60, with precision and recall for the least brightness level of -100 being 0.89 and 0.17 respectively. However, the rock maintains precision and recall in the range of 0.96 to 0.88 and 0.92 to 0.57 respectively. Background maintains the precision and recall in the range of 0.96 to 0.72 and 0.99 to 0.98 respectively.

It was observed that the effect of red channel reduction on model performance is significantly less compared to other noisy conditions. The model performance of $\mathcal{L}_{CL}$ and $\mathcal{L}_{JL}$ was observed without much fluctuation in the IoU score across various noise levels. However, $\mathcal{L}_{FL}$ and $\mathcal{L}_{DL}$ show some fluctuations in their performance.

7. Conclusion

This paper contributes to the underwater image analytics in two folds: (1) by creating novel annotated dataset for multiclass image segmentation for selected images from Fish4Knowledge dataset in coral reef scenarios; (2) by proposing a multiclass image segmentation model based on U-Net and residual modules for highly turbid underwater fish images in coral reef environment, and extensive analysis of the model performance with different loss functions in various simulated underwater noise scenarios.

Image analytics is crucial for underwater exploration, especially for monitoring various species in habitats with rich biodiversity, such as coral reefs. However, various artifacts such as low-resolution, haziness and turbidity environments make the analytics challenging. Moreover, datasets are very scarce in underwater environments with fish in coral reef environment, especially for multi-class semantic segmentation scenarios. The Fish4Knowledge dataset, primarily a binary segmentation data, is considered as the candidate data for the analytics. This research extends the dataset to multiclass scenarios by additionally creating annotations for the coral/rock information as the third class and it has been observed that the addition of the new class improved the segmentation of the class fish to a better extent. The models trained and performing well with this dataset can be used for improved underwater analytics.

The research also proposes a multiclass version of ResUnet architecture for multi-class semantic segmentation using the mentioned dataset annotations. Multiclass ResUnet was found to perform better when compared to the performance of Unet architecture and U-Net models with different encoder backbones utilising VGG19 and Inception v3 architectures. It is inferred that ResUnet based model showed the better performance as the identity connection ensures the previous stage information is also conveyed to the successive stages of feature extraction, along with the effective conveyance of gradient flow while optimisation, compared to the other architectures. Multiclass ResUnet architecture is further optimized using four different loss functions such as Categorical Cross-Entropy, Categorical Focal loss, Dice loss, and Jaccard loss, with a special focus on fish. Various underwater noisy conditions are simulated in the test images and analysed for the robustness of the model. The model optimised with Jaccard loss ($\mathcal{L}_{JL}$) was found to be robust in all the cases. $\mathcal{L}_{JL}$ outperforms with fewer outliers and demonstrates better performance in identifying the primary target of interest fish with IoU score of 0.8545 and a mean IoU of 0.9575. $\mathcal{L}_{JL}$ was also found to be robust in various levels of blurriness, brightness and red channel suppressions. $\mathcal{L}_{JL}$ was able to extract the fish details with an IoU score of 0.8 till image blurriness corresponding to the image quality of SSIM 0.5, with an IoU score of 0.9 till the brightness level of -60 and maintains the IoU performance above 0.935 across all the red channel levels. Robustness showed by Jaccard loss function may be due to the region based optimisation advantage, comparing to the other loss functions that rely more on the pixel-based strategy, which helps in maintaining its performance even in the harsh underwater noisy scenarios.

The proposed work is limited to the multiclass dataset annotation on selected images for Plectroglyphido don dicky class from the fish4knowledge and may not be generalised for other classes. Moreover, the work considers only the four loss functions as Cross-Entropy, Focal, Dice and Jaccard losses for analysing the underwater noise scenarios and lacks loss functions that can consider the boundary differences of the predicted mask from ground truth.

In future, the proposed multiclass annotation can be extended for other 23 classes of fish species from the fish4knowledge dataset. Architectures and loss functions that can accommodate geometric information of targets and underwater noises can be designed to make better segmentation models for the fish species. Further, the model can be evaluated with the real-world data and fine-tuned for the unseen noise scenarios in the original training dataset.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments

The authors would like to acknowledge the support of the Naval Research Board, Govt. of India for providing financial assistance for this research.

References

Abdullah, F., & Jalal, A. (2023). Semantic segmentation based crowd tracking and anomaly detection via neuro-fuzzy classifier in smart surveillance system. *Arabian Journal for Science and Engineering*, *48*(2), 2173-2190. https://doi.org/10.1007/s13369-022-07092-x.

Anantharajah, K., Ge, Z.Y., McCool, C., Denman, S., Fookes, C., Corke, P., Tjondronegoro, D., & Sridharan, S. (2014). Local inter-session variability modelling for object classification. In *IEEE Winter Conference on Applications of Computer Vision* (pp. 309-316). IEEE. Steamboat Springs, CO, USA. https://doi.org/10.1109/wacv.2014.6836084.

Apeer. (2023). Automated image analysis: a scalable solution for reproducible results. *Arivis Cloud*. https://www.apeer.com.

Apprill, A., Girdhar, Y., Mooney, T.A., Hansel, C.M., Long, M.H., Liu, Y., Zhang, W.G., Kapit, J., Hughen, K., Coogan, J., & Greene, A. (2023). Toward a new era of coral reef monitoring. *Environmental Science & Technology*, *57*(13), 5117-5124. https://doi.org/10.1021/acs.est.2c05369.

Asner, G.P., Vaughn, N.R., Balzotti, C., Brodrick, P.G., & Heckler, J. (2020). High-resolution reef bathymetry and coral habitat complexity from airborne imaging spectroscopy. *Remote Sensing*, *12*(2), 310. https://doi.org/10.3390/rs12020310.

Asner, G.P., Vaughn, N.R., Martin, R.E., Foo, S.A., Heckler, J., Neilson, B.J., & Gove, J.M. (2022). Mapped coral mortality and refugia in an archipelago-scale marine heat wave. *Proceedings of the National Academy of Sciences*, *119*(19), e2123331119. https://doi.org/10.1073/pnas.2123331119.

Bhagat, P.K., & Choudhary, P. (2018). Image annotation: then and now. *Image and Vision Computing*, *80*, 1-23. https://doi.org/10.1016/j.imavis.2018.09.017.

Boom, B.J., Huang, P.X., He, J., & Fisher, R.B. (2012b). Supporting ground-truth annotation of image datasets using clustering. In *Proceedings of the 21st International Conference on Pattern Recognition* (pp. 1542-1545). IEEE. Tsukuba, Japan.

Boom, B., Huang, P., Beyan, C., Spampinato, C., Palazzo, S., He, J., Beauxis-Aussalet, E., Lin, S.I., Chou, H.M., Nadarajan, G., Chen-Burger, Y.H., van Ossenbruggen, J., Giordano, D., Hardman, L., Lin, F.P., & Fisher, R.B. (2012a). Long-term underwater camera surveillance for monitoring and analysis of fish populations. In *International Workshop on Visual Observation and Analysis of Animal and Insect Behavior (VAIB), in*

Conjunction with the 21st International Conference on Pattern Recognition (pp. 1-4). IAPR. New York.

Brunet, D., Vrscay, E.R., & Wang, Z. (2012). On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing, 21*(4), 1488-1499. https://doi.org/10.1109/tip.2011.2173206.

Cael, B.B., Bisson, K., Boss, E., Dutkiewicz, S., & Henson, S. (2023). Global climate-change trends detected in indicators of ocean ecology. *Nature, 619*(7970), 551-554. https://doi.org/10.1038/s41586-023-06321-z.

Chicchon, M., Bedon, H., Del-Blanco, C.R., & Sipiran, I. (2023). Semantic segmentation of fish and underwater environments using deep convolutional neural networks and learned active contours. *IEEE Access, 11*, 33652-33665. https://doi.org/10.1109/access.2023.3262649.

Cui, S., Zhou, Y., Wang, Y., & Zhai, L. (2020). Fish detection using deep learning. *Applied Computational Intelligence and Soft Computing, 2020*(1), 1-13. https://doi.org/10.1155/2020/3738108.

Duntley, S.Q. (1963). Light in the sea. *Journal of the Optical Society of America, 53*(2), 214. https://doi.org/10.1364/josa.53.000214.

Galdran, A., Pardo, D., Picón, A., & Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation, 26*, 132-145. https://doi.org/10.1016/j.jvcir.2014.11.006.

Ghosh, S., Das, N., Das, I., & Maulik, U. (2019). Understanding deep learning techniques for image segmentation. *ACM Computing Surveys, 52*(4), 1-35. https://doi.org/10.1145/3329784.

Giles, A.B., Ren, K., Davies, J.E., Abrego, D., & Kelaher, B. (2023). Combining drones and deep learning to automate coral reef assessment with RGB imagery. *Remote Sensing, 15*(9), 2238. https://doi.org/10.3390/rs15092238.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 Conference on Computer Vision and Pattern Recognition* (pp. 770-778). IEEE. Las Vegas, NV, USA. https://doi.org/10.1109/cvpr.2016.90.

He, Z., Cao, L., Luo, J., Xu, X., Tang, J., Xu, J., Xu, G., & Chen, Z. (2024). UISS-Net: underwater image semantic segmentation network for improving boundary segmentation accuracy of underwater images. *Aquaculture International, 32*(5), 5625-5638. https://doi.org/10.1007/s10499-024-01439-x.

Hoegh-Guldberg, O., Poloczanska, E.S., Skirving, W., & Dove, S. (2017). Coral reef ecosystems under climate change and ocean acidification. *Frontiers in Marine Science, 4*(158), 1-20. https://doi.org/10.3389/fmars.2017.00158.

Horoszowski-Fridman, Y.B., Izhaki, I., Katz, S.M., Barkan, R., & Rinkevich, B. (2024). Shifting reef restoration focus from coral survivorship to biodiversity using reef carpets. *Communications Biology, 7*(1), 141. https://doi.org/10.1038/s42003-024-05831-4.

Islam, M.J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S.S., & Sattar, J. (2020). Semantic segmentation of underwater imagery: Dataset and benchmark. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1769-1776). IEEE. Las Vegas, NV, USA. https://doi.org/10.1109/iros45743.2020.9340821.

Jackett, C., Althaus, F., Maguire, K., Farazi, M., Scoulding, B., Untiedt, C., Ryan, T., Shanks, P., Brodie, P., & Williams, A. (2023). A benthic substrate classification method for seabed images using deep learning: application to management of deep-sea coral reefs. *Journal of Applied Ecology, 60*(7), 1254-1273. https://doi.org/10.1111/1365-2664.14408.

Ji, Y., Zhang, H., Zhang, Z., & Liu, M. (2021). CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. *Information Sciences, 546*, 835-857.

Khaled, A., Han, J.J., Ghaleb, T.A., & Mohamed, R. (2023). Fully convolutional neural network for improved brain segmentation. *Arabian Journal for Science and Engineering, 48*(2), 2133-2146. https://doi.org/10.1007/s13369-022-07169-7.

Kingma, D.P., & Ba, J. (2014). Adam: a method for stochastic optimization. *Computer Science, arXiv (Cornell*

University). https://doi.org/10.48550/arxiv.1412.6980.

Li, M., Zhang, H., Gruen, A., & Li, D. (2024). A survey on underwater coral image segmentation based on deep learning. *Geo-spatial Information Science*, 1-25. https://doi.org/10.1080/10095020.2024.2343323. (In press).

Li, X., Shang, M., Qin, H., & Chen, L. (2015). Fast accurate fish detection and recognition of underwater images with fast R-CNN. In *OCEANS 2015-MTS/IEEE Washington* (pp. 1-5). IEEE. Washington, DC. https://doi.org/10.23919/oceans.2015.7404464.

Lyons, M.B., Murray, N.J., Kennedy, E.V., Kovacs, E.M., Castro-Sanguino, C., Phinn, S.R., Acevedo, R.B., Alvarez, A.O., Say, C., Tudman, P., Markey, K., Roe, M., Canto, R.F., Fox, H.E., Bambic, B., Lieb, Z., Asner, G.P., Martin, P.M., Knapp, D.E., Li, J., Skone, M., Goldenberg, E., Larsen, K., & Roelfsema, C.M. (2024). New global area estimates for coral reefs from high-resolution mapping. *Cell Reports Sustainability, 1*(2), 1-13. https://doi.org/10.1016/j.crsus.2024.100015.

Mentzel, S., Nathan, R., Noyes, P., Brix, K.V., Moe, S.J., Rohr, J.R., Verheyen, J., Van Den Brink, P.J., & Stauber, J. (2024). Evaluating the effects of climate change and chemical, physical, and biological stressors on nearshore coral reefs: a case study in the great barrier reef, Australia. *Integrated Environmental Assessment and Management, 20*(2), 401-418. https://doi.org/10.1002/ieam.4871.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(7), 3523-3542. https://doi.org/10.1109/tpami.2021.3059968.

Mizuno, K., Terayama, K., Hagino, S., Tabeta, S., Sakamoto, S., Ogawa, T., Sugimoto, K., & Fukami, H. (2020). An efficient coral survey method based on a large-scale 3-D structure model obtained by speedy sea scanner and U-Net segmentation. *Scientific Reports, 10*(1), 12416. https://doi.org/10.1038/s41598-020-69400-5.

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

Nezla, N.A., Haridas, T.M., & Supriya, M.H. (2021). Semantic segmentation of underwater images using U-Net architecture based deep convolutional encoder decoder model. In *2021 7th International Conference on Advanced Computing and Communication Systems* (Vol. 1, pp. 28-33). IEEE. Coimbatore, India. https://doi.org/10.1109/icaccs51430.2021.9441804.

Ortiz, M., & Hermosillo-Núñez, B. (2024). Quantifying stability and resilience of eco-social keystone species complexes for coastal marine ecosystems of the Caribbean Sea and eastern Pacific: applications in conservation and monitoring programmes. *Philosophical Transactions of the Royal Society B Biological Sciences, 379*(1909), 20230176. https://doi.org/10.1098/rstb.2023.0176.

Peng, Y.T., & Cosman, P.C. (2017). Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing, 26*(4), 1579-1594. https://doi.org/10.1109/tip.2017.2663846.

Rashid, A.R., & Chennu, A. (2020). A trillion coral reef colors: deeply annotated underwater hyperspectral images for automated classification and habitat mapping. *Data*, *5*(1), 19. https://doi.org/10.3390/data5010019.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Springer International Publishing, Cham, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.

Sakshi, & Kukreja, V. (2023). Image segmentation techniques: Statistical, comprehensive, semi-automated analysis and an application perspective analysis of mathematical expressions. *Archives of Computational Methods in Engineering, 30*(1), 457-495. https://doi.org/10.1007/s11831-022-09805-9.

Saleh, A., Laradji, I.H., Konovalov, D.A., Bradley, M., Vazquez, D., & Sheaves, M. (2020). A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports*, *10*(1), 14671. https://doi.org/10.1038/s41598-020-71639-x.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition.* https://doi.org/10.48550/arxiv.1409.1556.

Singh, N., & Bhat, A. (2023). A systematic review of the methodologies for the processing and enhancement of the underwater images. *Multimedia Tools and Applications*, *82*(25), 38371-38396. https://doi.org/10.1007/s11042-023-15156-9.

Song, H., Mehdi, S.R., Zhang, Y., Shentu, Y., Wan, Q., Wang, W., Raza, K., & Huang, H. (2021). Development of coral investigation system based on semantic segmentation of single-channel images. *Sensors*, *21*(5), 1848. https://doi.org/10.3390/s21051848.

Stierhoff, K., & Cutter, G. (2013). Rockfish training and validation image dataset. *NOAA Southwest Fisheries Science Center Remotely Operated Vehicle Digital Still Images.* https://www.st.nmfs.noaa.gov/aiasi/SWFSC_ROV.html.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826). IEEE. Las Vegas, NV, USA. https://doi.org/10.1109/cvpr.2016.308.

Thampi, L., Thomas, R., Kamal, S., Balakrishnan, A.A., Haridas, T.P.M., & Supriya, M.H. (2021, December). Analysis of U-Net based image segmentation model on underwater images of different species of fishes. In *2021 International Symposium on Ocean Technology (SYMPOL)* (pp. 1-5). IEEE. Kochi, India. https://doi.org/10.1109/sympol53555.2021.9689415.

Thomas, T., Maurya, P., Manikandan, B., & Dessai, N.B.F. (2022). Estimation of coral reef area through 2d images: deep learning way using U-Net. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4195235.

Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T., & Mouillot, D. (2016, October). Coral reef fish detection and recognition in underwater videos by supervised machine learning: comparison between deep learning and HOG+SVM methods. In: Blanc-Talon, J., Distante, C., Philips, W., Popescu, D., Scheunders, P. (eds) *International Conference on Advanced Concepts for Intelligent Vision Systems* (pp. 160-171). Cham, Springer International Publishing. ISBN: 978-3-319-48680-2. https://doi.org/10.1007/978-3-319-48680-2_15.

Wang, Z., Bovik, A.C., Sheikh, H.R., & Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600-612. https://doi.org/10.1109/tip.2003.819861.

Yadav, R., & Pandey, M. (2022). Image segmentation techniques: a survey. In: Gupta, D., Polkowski, Z., Khanna, A., Bhattacharyya, S., Castillo, O. (eds) *Proceedings of Data Analytics and Management.* Springer Nature, Singapore, pp. 231-239. https://doi.org/10.1007/978-981-16-6289-8_20.

Yakut, C., Oksuz, I., & Ulukaya, S. (2023). A hybrid fusion method combining spatial image filtering with parallel channel network for retinal vessel segmentation. *Arabian Journal for Science and Engineering*, *48*(5), 6149-6162. https://doi.org/10.1007/s13369-022-07311-5.

Yang, L., Chen, Y., Shen, T., & Li, D. (2023). An FSFS-Net method for occluded and aggregated fish segmentation from fish school feeding images. *Applied Sciences*, *13*(10), 6235. https://doi.org/10.3390/app13106235.

Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., Yang, T. & Gao, M. (2023). Techniques and challenges of image segmentation: A review. *Electronics*, *12*(5), 1199. https://doi.org/10.3390/electronics12051199.

Yuan, M.-H., Lin, K.-T., Pan, S.-Y., & Yang, C.-K. (2024). Exploring coral reef benefits: A systematic SEEA-driven review. *Science of The Total Environment*, *950*, 175237. https://doi.org/10.1016/j.scitotenv.2024.175237.

Zhang, H., Li, M., Zhong, J., & Qin, J. (2024). CNet: A novel seabed coral reef image segmentation approach based on deep learning. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops* (pp. 767-775). IEEE. Waikoloa, HI, USA. https://doi.org/10.1109/wacvw60836.2024.00090.

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters, 15*(5), 749-753. https://doi.org/10.1109/lgrs.2018.2802944.

Zhong, J., Li, M., Zhang, H., & Qin, J. (2023). Fine-Grained 3D modeling and semantic mapping of coral reefs using photogrammetric computer vision and machine learning. *Sensors, 23*(15), 6753. https://doi.org/10.3390/s23156753.

Zhou, J., Yang, T., & Zhang, W. (2023). Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Applied Intelligence*, *53*(3), 3594-3621. https://doi.org/10.1007/s10489-022-03767-y.

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.