Autonomous Drone-Based Real-Time Maritime Search and Rescue Using Onboard CV and Precision Payload Deployment

Abstract

Maritime Search and Rescue (SAR) operations are inherently time-critical and demand rapid response over vast and often unpredictable oceanic environments. Traditional manual and semi-automated methods frequently fall short in delivering the speed, scalability, and accuracy required for effective rescues. This research presents an autonomous drone-based system equipped with an advanced computer vision model for enhancing SAR effectiveness through timely detection and intervention. The proposed system responds to emergency alarms, autonomously navigates to the incident location, hovers at a stable altitude of 200 meters, identifies individuals who have fallen into the water using an onboard vision model, and precisely deploys a life buoy to aid in rescue. To validate the detection component, we evaluated multiple variants of state-of-the-art object detection models, including different configurations of the YOLOv8 and YOLOv11 families trained on the SeaDronesSee dataset. Among the tested configurations, YOLOv8n and YOLOv11n demonstrated the best balance of accuracy and efficiency, achieving mAP_{50} scores of 0.621 and 0.611 respectively. Notably, YOLOv8n outperformed other variants in terms of F1 score (0.693), accuracy (0.492), and computational efficiency (8.1 GFLOPs), making it more suitable for real-time deployment on edge devices onboard drones. The study highlights the promising synergy between autonomous drone platforms and advanced machine vision, offering a scalable and resource-efficient solution for real-time victim detection and life-saving interventions in maritime environments.

Keywords: Maritime Search, Autonomous, Drones, YOLOv8, YOLOv11, Machine Vision, Rescue Operations, Precision Payload Deployment

1. Introduction

The open sea, vast, unpredictable, and often unforgiving, has long presented significant challenges to human safety. From shipwrecks to individuals accidentally falling overboard, maritime environments frequently become the setting for life-threatening emergencies. In such scenarios, Maritime Search and Rescue (SAR) operations play a critical role in saving lives. However, these operations are often hindered by the sheer scale of the ocean, unpredictable weather conditions, and the difficulty in locating victims promptly. Traditional SAR methods, whether manual or semi-automated, struggle to meet the urgency and precision demanded by these life-critical situations, underscoring the

need for more advanced, responsive, and efficient systems. Recent technological advancements offer promising alternatives.

Among these, Unmanned Aerial Vehicles (UAVs), commonly known as drones, stand out as transformative tools. Drones are flying machines that operate either autonomously or under remote control without onboard human presence. Based on their structural and functional characteristics, UAVs can be classified by size (mini, small, large) and are employed in diverse applications ranging from aerial photography and agriculture to infrastructure monitoring and emergency response. While large drones are often reserved for military applications, the increased accessibility and versatility of mini and small UAVs (weighing up to 25 kg) have accelerated their adoption in commercial and research domains.

In particular, autonomous drones, those capable of operating without constant human input, leverage onboard systems such as GPS, inertial measurement units (IMUs), and real-time control loops to navigate and execute complex tasks. These capabilities make them ideal for time-sensitive, high-risk applications like maritime SAR. Given the difficulty of accessing and monitoring vast oceanic regions, integrating drone technology can significantly improve the speed, precision, and effectiveness of victim detection and rescue response.

Furthermore, the integration of computer vision (CV) and artificial intelligence (AI) has elevated the capabilities of drones in autonomous operations. Object detection algorithms, especially real-time models like those in the YOLO (You Only Look Once) family, enable drones to identify and track people or objects in challenging marine environments. These models can be optimized for edge deployment, making real-time onboard inference feasible.

To evaluate these capabilities, we assessed a range of state-of-the-art object detection models from the YOLOv8 and YOLOv11 families, experimenting with multiple model sizes, such as nano (n), small (s), and medium (m), to determine the best balance between detection performance and real-time efficiency. Among these, YOLOv8n demonstrated better overall performance, achieving an F1 score of 0.693, accuracy of 0.429, and mAP_{50} of 0.621, compared to YOLOv11n's F1 score of 0.684, accuracy of 0.481, and mAP_{50} of 0.611. Despite YOLOv11n showing slightly lower computational complexity (6.3 GFLOPs), YOLOv8n's stronger F1 score and accuracy, along with its reasonable computational footprint (8.1 GFLOPs), made it more suited for real-time onboard deployment.

This research contributes to the growing field of autonomous drone-assisted SAR systems by developing a complete response pipeline: from the triggering of an alarm to drone deployment, real-time victim detection, and autonomous delivery of a life buoy to the victim's location. By combining machine vision and UAV technology, this work highlights a scalable and practical approach to improving real-time maritime rescue effectiveness. The findings from this study aim to inform future deployment strategies and help overcome limitations of current SAR operations.

1.1 Background Motivation

Maritime regions like the Mediterranean Sea have long faced recurring tragedies involving overcrowded vessels and migrants in distress. In one instance, 136 lives were lost in September 2012, followed by 339 deaths in October 2023, and a later incident where 34 victims drowned despite 150 being rescued by the Maltese Navy. These events highlight the urgent need for faster, more efficient maritime Search and Rescue (SAR) systems. Closer to home, India's 7,500 km coastline sees frequent maritime emergencies. Incidents such as the 2018 Cyclone Ockhi, where over 60 fishermen went missing, and the 2022 Mumbai cargo vessel capsize underline the limitations of current SAR methods. Coastal states like Goa, Kerala, and Tamil Nadu also report regular drowning cases due to strong currents and inadequate response time. Traditional SAR operations often suffer from slow deployment, limited visibility, and the vast scale of open waters. In critical rescue situations, delays cost lives. This research is driven by the need to enhance SAR capabilities through autonomous drone systems equipped with real-time object detection. By integrating YOLO-based machine vision models, specifically YOLOv8n and YOLOv11n, the proposed system aims to locate drowning victims and autonomously deliver life buoys. Among the two, YOLOv8n demonstrates superior overall performance, making it a more suitable choice for deployment in real-world maritime emergencies. The solution offers a scalable and rapid-response approach to improving the effectiveness of maritime rescue operations.

1.2 Dataset

The dataset employed in this study is a modified version of the SeaDronesSee dataset, originally developed by researchers at the University of Tübingen. It comprises 10,474 images categorized into five classes: boat, buoy, jetski, life_saving_appliances, and swimmer. Captured by 20 different subjects, the dataset features aerial footage of open waters under varying lighting conditions, offering a realistic training ground for object detection in maritime environments. To enhance the robustness of the model, the dataset was further enriched using cutout augmentation, a data augmentation technique that improves generalization by randomly masking square regions within the input images. For training purposes, the dataset was divided using a 70-20-10 split for training, validation, and testing, respectively, resulting in 7,322 training images, 2,077 validation images, and 1,075 test images. These figures are detailed in Table 1.

Table 1. Dataset Partitioning

Sub-set	Proportion	Images
Train	70%	7,322
Val	20%	2,077
Test	10%	1,075

1.3 YOLO Object Detection

YOLO (You Only Look Once) is a widely adopted algorithm renowned for its high-performance object detection capabilities. Since the introduction of the YOLO series in 2015, the framework has evolved through multiple versions, each offering notable improvements in speed and accuracy. In this study, we trained and evaluated multiple model sizes from the YOLOv8 and YOLOv11 families, including the n (nano), s (small), and m (medium) variants. The research involved extensive experimentation, including multiple rounds of parameter tuning, model scaling, and architectural adjustments, to identify an optimal model that delivers strong performance across evaluation metrics while remaining feasible for real-time deployment on drones. Among the tested variants, YOLOv8n demonstrated superior performance in terms of detection accuracy and overall balance between speed and resource efficiency.

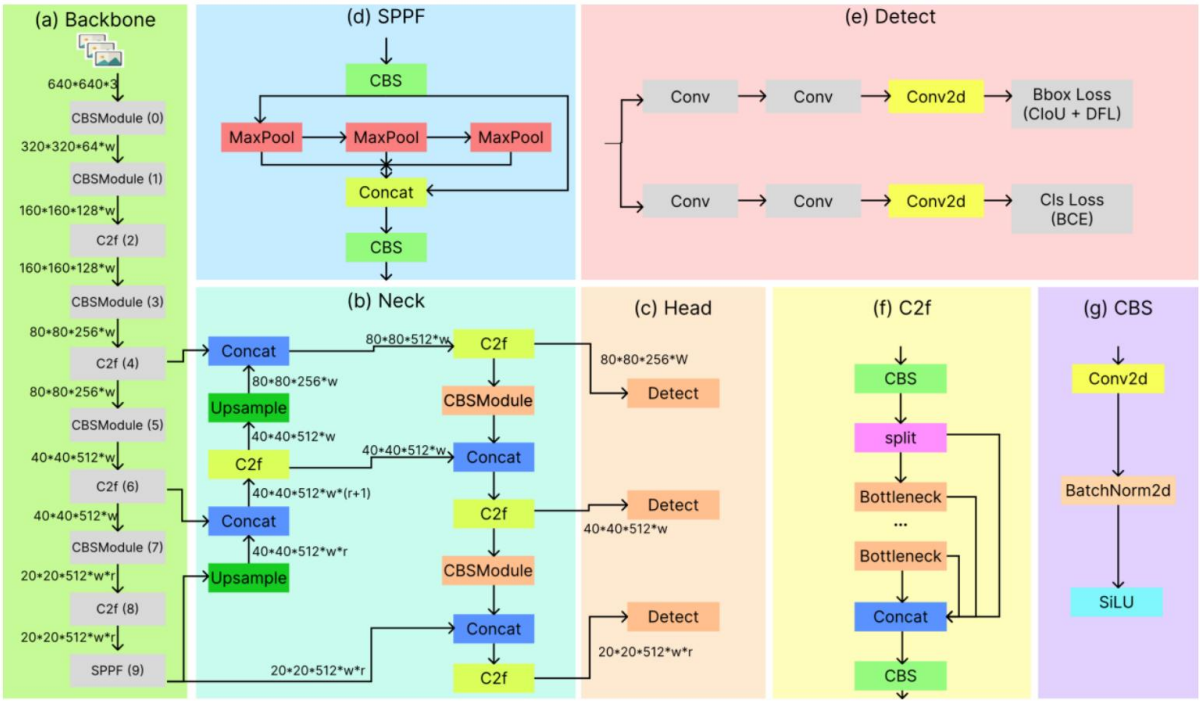

Figure 1. Schematic figure of the network structure of YOLOV8

YOLOv8 employs a modified CSPDarknet53 architecture as its backbone, progressively downsampling features across five stages (B1 through B5) to generate multi-scale feature maps. A key

innovation is the use of the lightweight C2f module, replacing the original Cross Stage Partial (CSP) module. The C2f module includes a gradient shunt connection that facilitates better information flow during feature extraction while maintaining computational efficiency.

Feature extraction in YOLOv8 is supported by the CBS module, which performs a sequence of convolution, batch normalization, and SiLU activation. At the end of the backbone, a Spatial Pyramid Pooling – Fast (SPPF) module condenses feature maps into fixed-size representations using three sequential max-pooling layers. This approach offers reduced latency compared to traditional SPP methods.

For feature aggregation, YOLOv8 uses a PAN-FPN neck, inspired by PANet. Unlike earlier versions such as YOLOv5 and YOLOv7, YOLOv8 omits convolution layers after upsampling, reducing architectural complexity without compromising performance. The dual-path PAN-FPN design combines features from both top-down and bottom-up flows, effectively merging deep semantic information with spatial detail to support accurate object localization.

The detection head in YOLOv8 uses a decoupled architecture with separate branches for classification and bounding box regression. Binary Cross Entropy (BCE) loss is used for classification, while a combination of Distribution Focal Loss (DFL) and Complete IoU (CIoU) loss is used for bounding box refinement. YOLOv8 also adopts an anchor-free approach and employs a Task-Aligned Assigner to dynamically determine positive and negative samples during training, thereby enhancing both accuracy and robustness.

Building upon the architecture and design principles of YOLOv8, YOLOv11 introduces further enhancements for detection and segmentation tasks. It features an optimized backbone and neck structure that improves feature extraction while maintaining efficiency and adaptability. In our comparative experiments, we evaluated YOLOv11n, YOLOv11s, and YOLOv11m. While YOLOv11n also delivered competitive results, YOLOv8n outperformed it overall in this study. YOLOv11 maintains a lightweight design and compatibility with edge and cloud platforms, making it suitable for a broad range of deployment scenarios. It supports advanced tasks including oriented object detection and segmentation, offering flexibility for high-performance, real-time applications.

2. Literature Survey

2.1 Object Detection in Rescue Operations

Recent advancements have seen the integration of drone-based object detection systems in various rescue scenarios. For instance, YOLOv4 has been effectively used to detect individuals involved in accidents during high-risk outdoor activities such as skiing, hiking, and mountain biking. To overcome the limitations of traditional vision-based systems in low-visibility conditions, Thermal Infrared (TIR) cameras have been employed for automatic human detection in search and rescue (SAR) missions.

Additionally, convolutional neural network (CNN) models have been developed to detect critical ground features from aerial imagery in post-disaster environments. These models, trained on the custom Volan2018 aerial video dataset, are capable of identifying damaged and intact rooftops, vehicles, vegetation, debris, and flood zones, demonstrating the effectiveness of CNNs in disaster response and assessment.

2.2 Summary of SeaDronesSee dataset in Rescue tasks

The SeaDronesSee dataset was created to address the lack of suitable datasets tailored for maritime search and rescue (SAR) operations. Previous datasets focused primarily on remote sensing using synthetic aperture radar (SAR) imagery, which relied on satellite-captured top-down views. While effective for detecting large vessels, these datasets are inadequate for identifying smaller objects such as swimmers. Additionally, satellite imagery is often hindered by environmental factors like cloud cover, reducing its reliability in time-critical rescue missions. In contrast, the SeaDronesSee dataset offers high-resolution RGB imagery, ranging from 3840 × 2160 px to 5456 × 3632 px, and includes dedicated object classes such as boats, jet skis, buoys, life-saving appliances, and swimmers, making it more suitable for close-range drone-based SAR tasks.

2.3 Object Detection models utilizing the SeaDronesSee dataset

The SeaDronesSee dataset was created to address the lack of suitable datasets tailored for maritime search and rescue (SAR) operations. Previous datasets focused primarily on remote sensing using synthetic aperture radar (SAR) imagery, which relied on satellite-captured top-down views. While effective for detecting large vessels, these datasets are inadequate for identifying smaller objects such as swimmers. Additionally, satellite imagery is often hindered by environmental factors like cloud cover, reducing its reliability in time-critical rescue missions. In contrast, the SeaDronesSee dataset offers high-resolution RGB imagery, ranging from 3840 × 2160 px to 5456 × 3632 px, and includes dedicated object classes such as boats, jet skis, buoys, life-saving appliances, and swimmers, making it more suitable for close-range drone-based SAR tasks.

3. Methodology

3.1 Dataset Preparation

A refined version of the SeaDronesSee dataset, originally comprising 5,630 annotated images, was selected for this study. These images were captured using five distinct camera systems (see Table 2) mounted on three different drones, DJI Matrice 100, DJI Matrice 210, DJI Mavic 2 Pro, and a fixed-wing Trinity F90+ aircraft developed by Quantum Systems. The diversity of imaging platforms was intended to minimize camera bias and ensure broader generalization across different visual conditions. The initial annotations were created using DarkLabel, a free and open-source labeling tool, and

classified into five categories: swimmer (person in water without a life jacket), floater (person in water with a life jacket), swimmer† (person on boat without a life jacket), floater† (person on boat with a life jacket), and boats.

Table 2. Specifications of cameras used in generating the SeaDronesSee dataset

Camera	Resolution	Purpose
Hasselblad L1D-20c	3840 * 2160	Video capture at 30 fps
MicaSense RedEdge-MX	1280 * 960	Multi-spectral capture at 1 fps
Sony UMC-R10C	5456 * 3632	Image capture
Zenmuse X5	3840 * 2160	Video capture at 30 fps
Zenmuse XT2	3840 * 2160	Video capture at 30 fps

For this research, we adopted the Roboflow SeaDronesSee v10 dataset, an augmented and reorganized version of the original. This version features 10,474 images, expanded through cutout augmentation and auto-orientation techniques. The annotation schema was restructured to define five new consolidated classes: boat, buoy, jetski, life_saving_appliances, and swimmer. Notably, the 'swimmer' class aggregates all four person-related categories from the original dataset. The class-wise distribution of these merged categories is shown in Table 3, and their instance frequencies in the training dataset are visualized in Figure 2.

Table 3. Frequency of class image across dataset images

Class	Images
Swimmer	8,185
Boat	6,782
Buoy	4,073
Jetski	2,648
Life_saving_appliances	856

3.2 Conceptual Framework for Autonomous Aerial Life Buoy Deployment

This research proposes a conceptual autonomous drone-based rescue system specifically tailored for maritime search and rescue (SAR) missions involving individuals who have fallen overboard or into nearshore waters. The envisioned pipeline integrates emergency detection, autonomous navigation, real-time object detection, and targeted payload deployment, forming a closed-loop rescue response system. While not physically implemented in this phase of the study, each component of the

system is designed to be grounded in practical deploy-ability and informed by current state-of-the-art computer vision and robotics frameworks.

i. Incident Trigger and Alarm Interface

The operational cycle begins with the triggering of an emergency alarm. This alert may originate from a variety of sources, including wearable sensors on passengers, ship-based fall-detection systems, coastal surveillance units, or visual verification by human observers. Upon confirmation, the alert is relayed to a central command module interfaced with a fleet of standby unmanned aerial vehicles (UAVs).

ii. Autonomous Drone Dispatch and Path Planning

Upon receiving the incident coordinates, a rescue UAV is immediately deployed. The drone is programmed to execute an autonomous take-off routine, followed by real-time path planning toward the alarm zone. The drone's navigation system leverages GPS and pre-mapped geospatial data to compute an optimal route, balancing speed and safety. The route planning logic incorporates several modules:

- **Waypoint Generation**: A set of dynamic waypoints are generated between the drone's base station and the target location, considering both geodetic distance and environmental constraints (e.g., no-fly zones, restricted airspace, or weather conditions).
- **Obstacle Avoidance**: Using onboard LiDAR, radar, or depth-sensing cameras (in future hardware integrations), the UAV can dynamically adjust its path in response to sudden obstructions such as terrain features, tall structures, or other aircraft.
- **Adaptive Altitude Control**: The UAV maintains an altitude of approximately 200–250 meters over water to ensure a wide field of view for visual detection while preserving spatial resolution for effective object recognition by the onboard camera.

The drone continues to the specified alarm region, hovering and loitering in a circular or grid-based search pattern if no swimmer is immediately detected upon arrival.

iii. Visual Search and Detection Using Onboard Computer Vision

Once over the search area, the UAV activates its downward-facing RGB camera system, calibrated to capture high-resolution imagery compatible with the inference requirements of the detection model. The onboard YOLOv8n object detection model, pretrained and fine-tuned on the SeaDronesSee v8 dataset, is deployed to scan each video frame in real time.

Key features of the detection phase include:

- **Frame-wise Inference:** The RGB video stream is processed frame-by-frame at 30 FPS, with bounding boxes, class labels, and confidence scores logged per frame.

- **Detection Validation:** To mitigate false positives due to water reflections, waves, or floating debris, the system uses temporal smoothing logic. A detection is confirmed only if the same object class (i.e., swimmer) appears with a confidence ≥ 0.7 in at least three consecutive frames.
- **Localization and Target Tracking:** Upon confirmation, the centroid coordinates of the bounding box are extracted and converted into geospatial coordinates using camera intrinsics and GPS-altitude metadata. The drone then switches to a target-following mode, continuously updating the swimmer's location as they drift or move with the waves.

Figure 2. Drone Navigating towards Stranded People over the Ocean

iv. Precision Payload Deployment

Once the swimmer's location is stabilized within a defined spatial tolerance (e.g., deviation < 2 meters across three frames), the UAV repositions directly over the centroid of the target. The drone transitions to hover mode, maintains positional lock using GPS and barometer fusion, and prepares for payload deployment. The payload mechanism is designed to release a single-use life-saving buoy.

Though this research phase does not implement the mechanical deployment, the complete logic for trajectory prediction and visual confirmation is established, forming a reliable basis for future integration with drop-capable UAV platforms.

v. Operational Evaluation Metrics

The performance of the system is intended to be evaluated based on three primary metrics:

- **Detection Accuracy:** Derived from the YOLOv11n model's swimmer-class precision, recall, and mAP metrics as benchmarked on SeaDronesSee v10.

- **Drop Precision:** Defined as the Euclidean distance between the predicted swimmer coordinates and the actual point of buoy impact in water.
- **System Latency:** Measured from the timestamp of swimmer entry in the visible field to the moment of detection confirmation and drop execution.

Figure 3. Drone Carrying Life Jacket Payload

This autonomous rescue pipeline emphasizes speed, accuracy, and real-time responsiveness, and is structured to operate effectively in dynamic, high-risk maritime environments. The methodology, while conceptual in this paper, offers a viable roadmap for future development and deployment of aerial life-saving systems powered by deep learning and autonomous robotics.

3.3 Model Configurations

The configurations implemented in this study were carefully designed to thoroughly evaluate the performance of the YOLOv8 and YOLOv11 models under a wide range of experimental conditions. Each model family, YOLOv8 and YOLOv11, was explored across multiple scale variants, specifically the n (nano), s (small), and m (medium) versions. These variants were tested using various input image sizes, anchor settings, and hyperparameter combinations to identify the most effective setup for maximizing detection accuracy while minimizing computational cost.

Particular attention was given to tuning critical hyperparameters such as learning rate, batch size, and number of training epochs to ensure each model achieved both precision and efficiency. YOLOv8 was optimized to maintain a balanced trade-off between speed and accuracy, while YOLOv11 incorporated architectural advancements, such as refined backbone structures and enhanced feature extraction modules, to further boost detection capability and robustness across diverse scenarios.

After thorough evaluation, the models ultimately selected for final training and integration into the drone pipeline were YOLOv8n.pt and YOLOv11n.pt. These variants demonstrated the best overall

balance of detection performance and real-time inference efficiency, making them ideal for deployment on resource-constrained edge devices onboard UAVs. While s and m variants showed marginally better accuracy, their significantly higher computational requirements (e.g., GFLOPs, memory, latency) made them less viable for practical use in maritime SAR missions.

Given that SAR operations often involve identifying small, dispersed, and low-resolution targets, such as swimmers or floating life-saving appliances, additional network adaptations were incorporated to improve localization. Specifically, a high-resolution P2 detection layer was added to the model architecture via the .yaml configuration file. This layer increased spatial granularity in early feature maps, enhancing small object detection. The .yaml file also defined consistent train, validation, and test dataset paths to ensure standardized training across all variants.

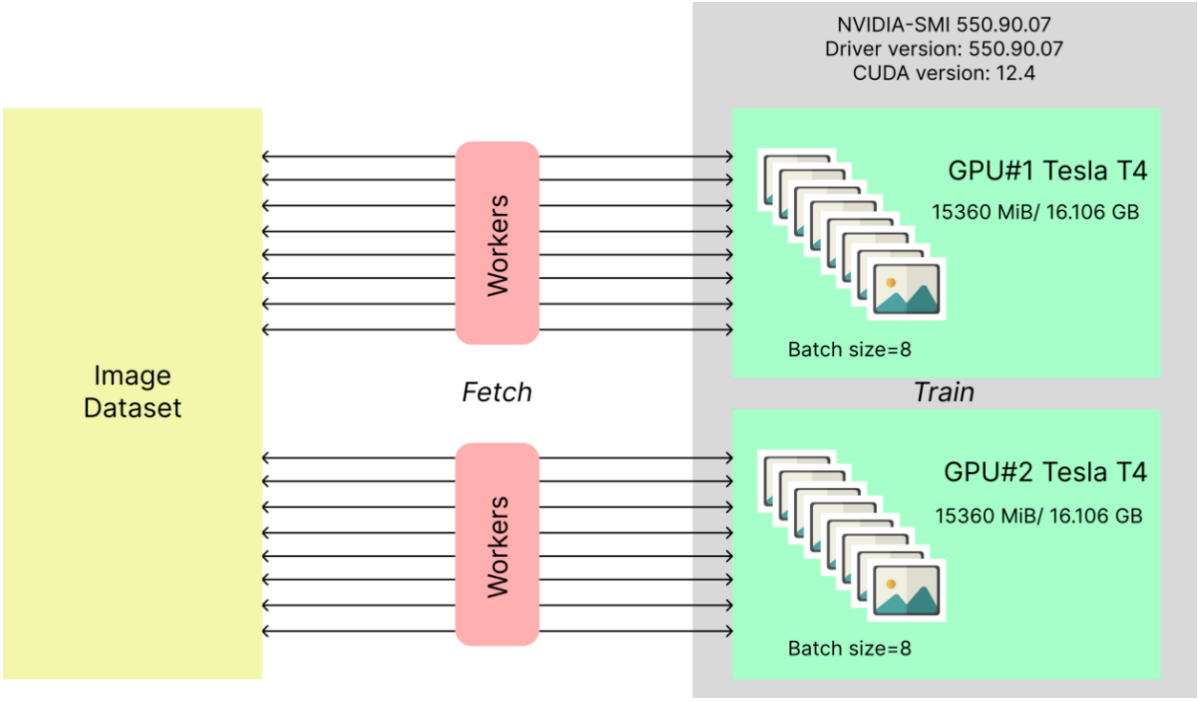

Figure 3. NVIDIA-SMI 550.90.07

Model training was conducted using the Kaggle platform, which provides access to advanced GPU infrastructure. For this study, Kaggle's compute environment offered limited-time usage of NVIDIA-SMI version 550.90.07, featuring a dual-GPU setup with Tesla T4 GPUs. As illustrated in Figure 3, the environment supports efficient multi-threaded data loading using parallel workers. These workers fetch and preprocess image batches, minimizing I/O delays and ensuring steady GPU utilization. The dataset was then distributed in parallel to the two GPUs, which enabled faster model training and reduced time-to-convergence due to distributed computation.

All models were trained for a total of 100 epochs, which was found to be an optimal point for balancing model performance while mitigating risks of overfitting. Both the batch size and the number of workers were set to 8, aligning with the memory constraints of the GPU environment while also speeding up the data processing pipeline. To improve detection in crowded maritime scenes, an Intersection over Union (IoU) threshold of 0.7 was employed. This high threshold allowed for more accurate object distinction by reducing the overlap in predicted bounding boxes. The training leveraged the AdamW optimizer with the momentum setting kept to 'Auto', which dynamically adjusted during training. The final models, YOLOv8n and YOLOv11n, used a momentum value of 0.9 and a learning rate of 0.000714, selected based on iterative experimentation for stable and effective convergence across both model variants.

3.4 Evaluation Metrics

Across various annotated datasets employed by object detection challenges and the research community, the primary metric for evaluating detection accuracy is Average Precision (AP). To fully understand the variations of AP, it is essential to first familiarize ourselves with some fundamental terms commonly used in this context:

- **True Positive (TP):** A correct identification of a ground-truth bounding box.
- **False Positive (FP):** An incorrect identification, either detecting a non-existent object or misplacing the detection of an existing one.
- **False Negative (FN):** A failure to detect an actual ground-truth bounding box.

The count of True Negatives (TN) is irrelevant in object detection since the number of potential bounding that should not be identified within an image is infinite.

Precision and recall are important measures in machine learning that evaluate the performance of a model. Precision computes the correctness of positive predictions, representing the percentage of correct positive predictions. While recall determines how well the model recognizes all relevant instances.

A precision-recall curve (PR curve) is used for visualizing the relationship between Precision and Recall, across varying confidence thresholds assigned to the bounding boxes predicted by the detector.

In the 11-point interpolation method, the precision-recall curve is summarized by averaging the highest precision values at 11 evenly spaced recall levels: 0, 0.1, 0.2, ..., 1. The AP is calculated by considering the maximum interpolated precision, Pinterp(R) at each recall level, rather than the observed precision, P(R).

$$AP_{all} = \sum_{n} (R_{n+1} - R_n) P_{interp}(R_{n+1})$$

where,

$$P_{interp}(R_{n+1}) = \max_{\tilde{R}:\tilde{R}\geq R_{n+1}} P(\tilde{R})$$

And $\tilde{R}$ denotes the mean of R values.

To measure the accuracy over all classes, the average of AP over all classes is taken, this metric is known as the mean average precision (mAP.)

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

The average precision (AP) is a per-class measure while mAP is the average of APs across all the classes. Thus, mAP is a robust evaluation metric that considers multiple queries.

A confusion matrix is a tabular representation of a model's performance. It compares the predicted labels to the actual labels and provides detailed insights into the model's performance in each class. For object detection, the confusion matrix is typically used for evaluating the classification of detected objects. Figure: 7 shows the structure of a confusion matrix.

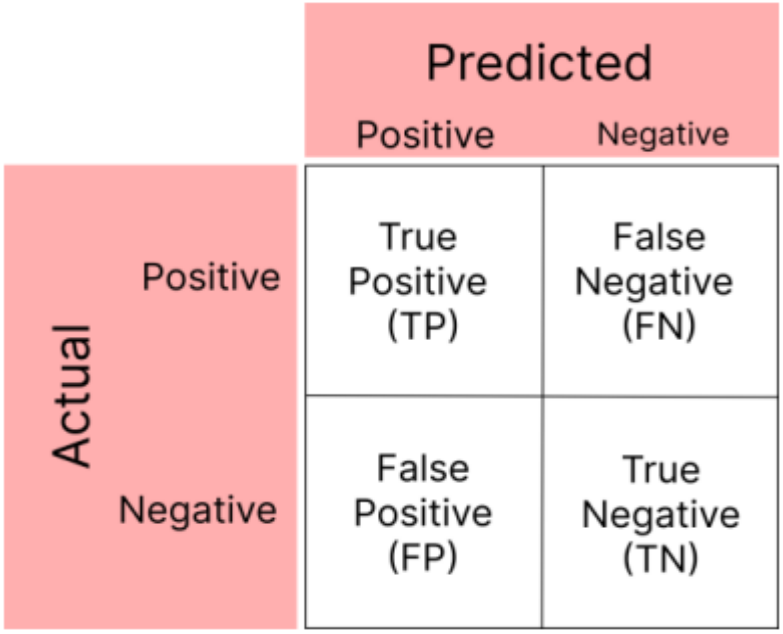

Figure 7. Structure of a confusion matrix

The structure of a confusion matrix has been depicted in Figure 7, which serves as an important evaluation tool in classification tasks. It consists of four key components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The matrix helps analyse model performance, providing insights into accuracy, precision, recall, and overall error rates for the trained model.

Formulae of evaluation metrics like Precision, Recall, Negative Predictive Value, Specificity, and Accuracy, are as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Negative\ Predicted\ Value = \frac{TN}{TN + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

The F1 score is a metric used to measure the performance of a machine learning model. It can be calculated by combining the precision and recall of the model. The F1 score can be calculated in the following way:

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$$

4. Result and Discussion

4.1 Quantitative Performance Analysis

Table 4. Performance comparison of the trained models across all the classes (scale of 0-1)

Model	GFLOPs	Precision	Recall	mAP_{50}	mAP_{50-95}	F1 score	Accuracy
YOLOv8n	8.1	0.818	0.601	0.621	0.393	0.693	0.492
YOLOv8s	28.4	0.845	0.633	0.660	0.408	0.724	0.558
YOLOv8m	78.7	0.691	0.638	0.666	0.412	0.663	0.571
YOLOv11n	6.3	0.810	0.592	0.611	0.369	0.684	0.481
YOLOv11s	21.3	0.855	0.628	0.668	0.410	0.724	0.546
YOLOv11m	67.7	0.848	0.632	0.692	0.422	0.724	0.571

In analysing the performance metrics across the YOLOv8 and YOLOv11 models, several key insights emerge, which have been documented in Table 4. YOLOv8 models span from 8.1 GFLOPs for YOLOv8n to 257.4 GFLOPs for YOLOv8x, while YOLOv11 models are generally more computationally efficient, ranging from 6.3 GFLOPs for YOLOv11n to 194.4 GFLOPs for YOLOv11x. Smaller models in both series consume significantly fewer computational resources, with YOLOv11 models typically requiring less. In terms of precision, YOLOv11 models slightly outperform YOLOv8 models, varying from 0.691 for YOLOv8m to 0.855 for YOLOv11s. The most precise models are YOLOv11s (0.855), YOLOv11m (0.848), and YOLOv8s (0.845), showcasing the high detection confidence of smaller variants in both series.

The recall is relatively level across models, ranging from a minimum of 0.592 for YOLOv11n to a maximum of 0.638 for YOLOv8m, with most models falling in the range of 0.6 to 0.64, revealing similar object detection capabilities. For mAP_{50}, YOLOv11m leads at 0.692, followed by YOLOv11s (0.668) and YOLOv8m (0.666), while for $mAP_{50\text{-}95}$, values range from 0.369 for YOLOv11n to 0.422 for YOLOv11m, indicating performance drop-off at stricter IoU thresholds. The F1 scores, which balance precision and recall, range between 0.663 for YOLOv8m and 0.724 for both YOLOv8s and YOLOv11m, indicating a strong balance in detection quality. Accuracy values span from 0.481 for YOLOv11n to 0.571 for YOLOv11m, with most models in the 0.54–0.57 range.

However, when balancing all metrics with real-time deployment constraints, YOLOv8n emerges as the most optimized model. Despite its modest computational requirement of just 8.1 GFLOPs, it delivers a strong F1 score of 0.693 and an accuracy of 0.492, surpassing its YOLOv11n counterpart (F1: 0.684, Accuracy: 0.481) while maintaining competitive values in precision (0.818), recall (0.601), and mAP_{50} (0.621). These characteristics make YOLOv8n particularly well-suited for edge deployment onboard UAVs in time-critical maritime search and rescue scenarios. In summary, although YOLOv11 models offer higher precision and slightly better computational efficiency, YOLOv8n provides the best overall trade-off between speed, accuracy, and resource usage, making it the preferred model for real-world SAR applications involving real-time victim detection and response.

4.2 Comparative mAP_{50} and $mAP_{50\text{-}95}$ Analysis

Figure 8 presents a comparison of mAP_{50} (mean average precision at IoU threshold 0.5) and $mAP_{50\text{-}95}$ (mean average precision across IoU thresholds 0.5 to 0.95) for the YOLOv8 series: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. As expected, both metrics generally improve with increasing model complexity, with YOLOv8x achieving the highest values, mAP_{50}: 0.687 and $mAP_{50\text{-}95}$: 0.429, reflecting its ability to maintain precision across stricter IoU thresholds. However, for real-time, resource-constrained deployments such as drone-assisted search and rescue, YOLOv8n stands out as

the most optimal choice. Despite its minimal computational footprint (8.1 GFLOPs), it delivers competitive mAP_{50-95} (0.393) and strong overall performance, striking an effective balance between accuracy and efficiency. This reinforces the importance of model selection tailored to the deployment context, where smaller models like YOLOv8n offer superior practical utility.

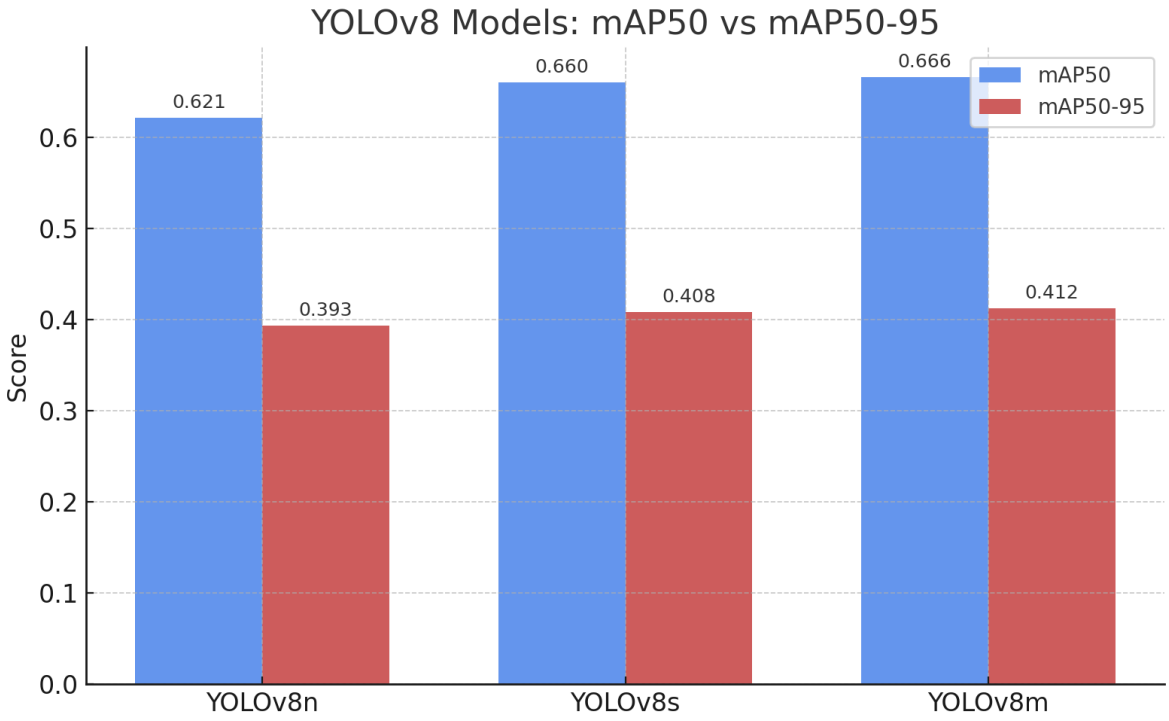

Figure 8. Column chart for comparing mAP_{50} and mAP_{50-95}values of YOLOv8 models

The mAP_{50} and mAP_{50-95} values for the various YOLOv11 models, nano, small, and medium, are illustrated in Figure 9. Unlike the consistent trend observed in the YOLOv8 series, YOLOv11 models exhibit a non-linear performance pattern. The mAP_{50} peaks at 0.692 with the YOLOv11m (medium) model, but does not show proportional gains in mAP_{50-95} or accuracy, and the performance diminishes for lighter models like YOLOv11n (mAP_{50}: 0.611, mAP_{50-95}: 0.369).

While YOLOv11m delivers the highest precision within its family, its higher GFLOPs (67.7) and only marginal gains over YOLOv8n make it less ideal for real-time deployment in constrained maritime environments.

This reinforces the conclusion that YOLOv8n, with its optimal balance of accuracy, efficiency, and general detection performance, remains the most practical and scalable choice for autonomous search and rescue systems.

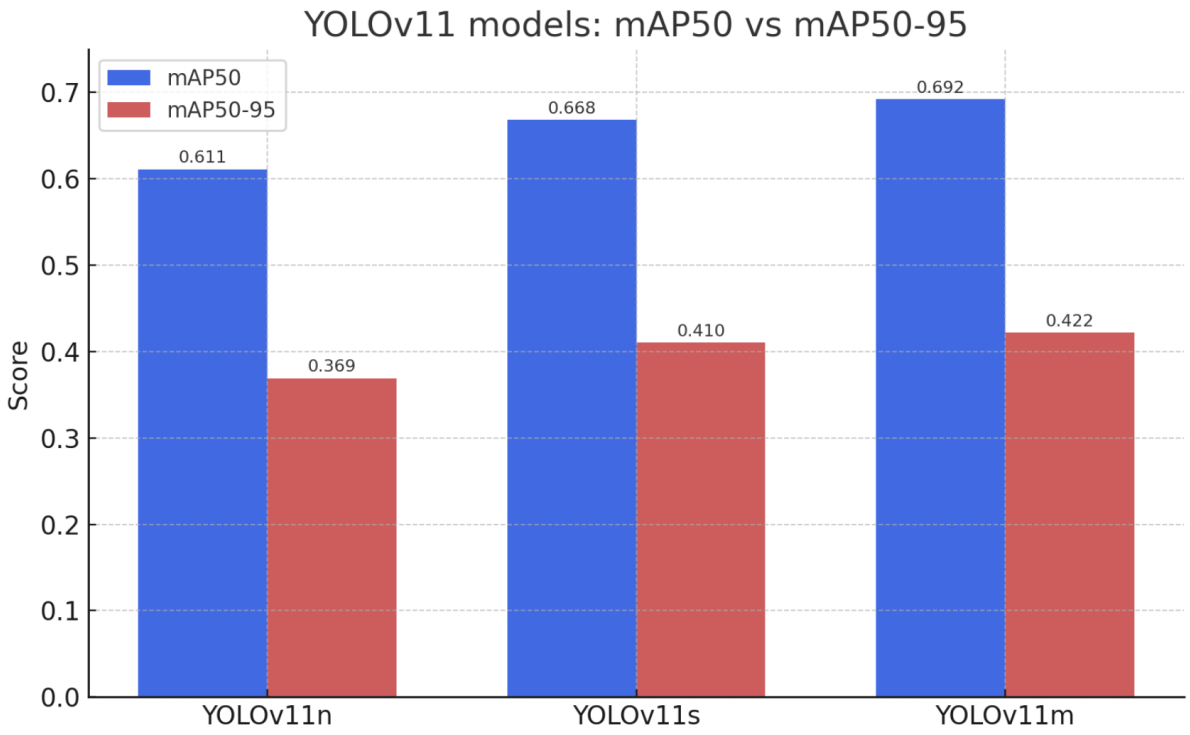

Figure 9. Column chart for comparing mAP_{50} and $mAP_{50\text{-}95}$values of YOLOv11 models

4.2 Comparative GFLOPs Analysis

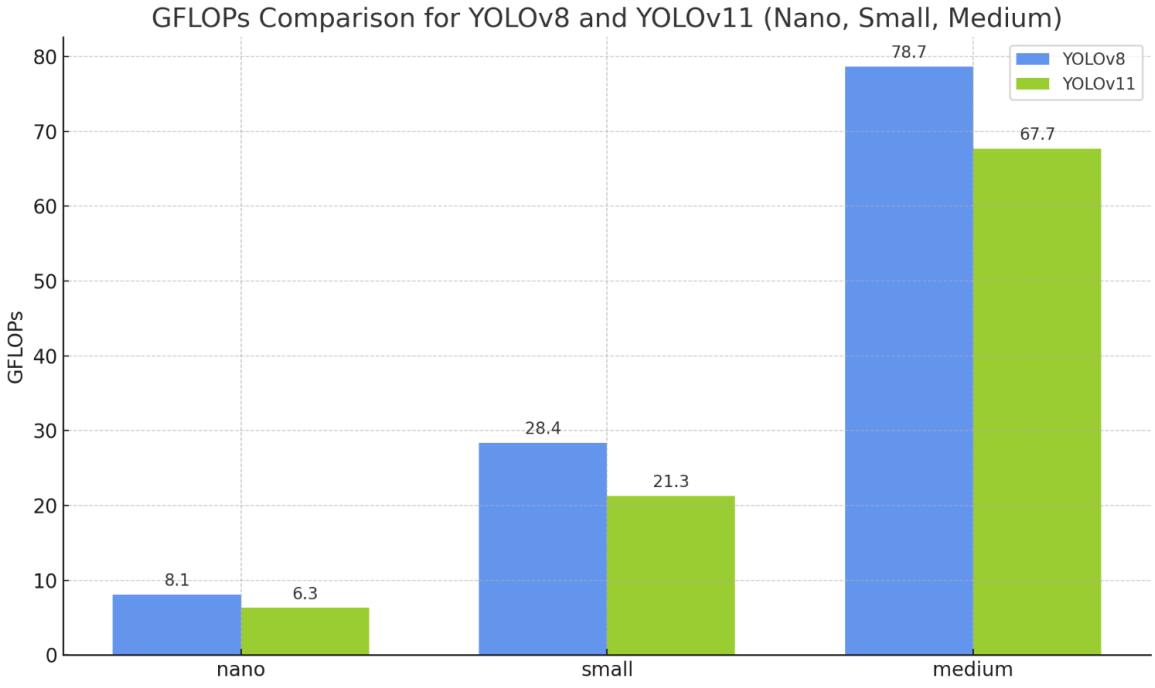

Figure 10. Comparing GFLOP values for different models across YOLOv8 and YOLOv11 series

A detailed comparison of GFLOPs (Giga Floating Point Operations) across the YOLOv8 and YOLOv11 model series is presented in Figure 10, offering insights into each model's computational demands. While YOLOv11 models, particularly in their smaller configurations, exhibit lower GFLOP values and are computationally lightweight, the YOLOv8 series, especially YOLOv8n, achieves a more optimal balance between accuracy and efficiency. YOLOv8n, despite a modest increase in computational load (8.1 GFLOPs vs. YOLOv11n's 6.3 GFLOPs), demonstrates superior accuracy, F1 score, and overall detection performance, making it the more suitable choice for edge-based, real-time maritime SAR applications. This comparison is critical for selecting models that meet both performance requirements and the constraints of onboard drone deployment.

4.3 Class-wise Performance Analysis of Top 2 performing models

The class-wise performances of YOLOv8n (Accuracy: 0.492) and YOLOv11n (Accuracy: 0.481) are tabulated in Table 5 (for YOLOv8n) and Table 6 (for YOLOv11n):

Table 5. Class-wise performance for YOLOv8n

Class	Precision	Recall	mAP_{50}	mAP_{50-95}
boat	0.927	0.819	0.857	0.615
buoy	0.935	0.715	0.738	0.448
jetski	0.924	0.893	0.875	0.595
life_saving_appliances	0.292	0.006	0.007	0.021
swimmer	0.740	0.560	0.574	0.213

Table 6. Class-wise performance for YOLOv11n

Class	Precision	Recall	mAP_{50}	mAP_{50-95}
boat	0.881	0.874	0.835	0.642
buoy	0.829	0.731	0.724	0.489
jetski	0.733	0.849	0.820	0.545
life_saving_appliances	0.963	0.000	0.162	0.059
swimmer	0.681	0.548	0.563	0.229

The poor performance of the 'life_saving_appliances' class as shown by Table 5 and Table 6 can be understood by Table 8, which shows the number of instances for every class in the validation dataset. The limited representation of the class in the dataset might result in its underrepresentation in the validation set, leading the model's struggle to generalize detection.

Table 7. Class-wise representation in the validation dataset

Class	Images	Instances
boat	1209	2413
buoy	806	934
jetski	768	769
life_saving_appliances	592	884
swimmer	1902	11151

An in-depth class-wise analysis of the YOLOv8n and YOLOv11n models places particular emphasis on the swimmer class, which is central to maritime search and rescue (SAR) operations. Detecting swimmers is inherently challenging due to factors such as partial submersion, varied postures, and relatively small visible areas. Despite these difficulties, both models exhibit moderate but promising performance. YOLOv8n achieves a precision of 0.740, recall of 0.560, mAP_{50} of 0.574, and mAP_{50-95} of 0.213, indicating a reasonably balanced ability to both identify and localize swimmers. YOLOv11n, while slightly lagging in swimmer precision (0.681) and recall (0.548), yields a comparable mAP_{50-95} of 0.229, suggesting that while its localization across IoU thresholds is acceptable, its detection reliability in real-world conditions may be slightly lower.

Larger and visually distinct objects such as boats and jetskis are detected with high accuracy by both models. YOLOv8n reports mAP_{50} values of 0.857 for boats and 0.875 for jetskis, while YOLOv11n closely follows with 0.835 and 0.820, respectively. Buoys also show consistent detectability, with mAP_{50} values above 0.72 for both models, supported by a reasonable number of labeled instances.

A stark contrast is observed in the life_saving_appliances class, where performance significantly deteriorates. Despite having 884 labelled instances across 592 images, YOLOv8n yields a very low precision of 0.292 and recall of 0.006, while YOLOv11n, though reporting perfect precision (0.963), exhibits zero recall. This discrepancy highlights a critical issue: although the models may confidently classify an object when detected, they almost never identify life-saving appliances in the first place. This failure can be attributed to several factors:

- **Small object size:** Life-saving appliances often occupy minimal pixel area, making them difficult to detect, especially for lightweight models with limited spatial resolution in early layers.
- **Limited representation in the validation set:** While there are 884 labelled instances in total, their distribution may be uneven across subsets (train/val/test), and if underrepresented during validation, the models struggle to generalize detection.
- **Visual similarity with background or occlusion:** These objects may blend into complex maritime scenes or be partially hidden, further complicating detection.

In conclusion, while both YOLOv8n and YOLOv11n are computationally efficient and suitable for real-time deployment, YOLOv8n demonstrates greater reliability for swimmer detection, making it more aligned with the primary objective of SAR-focused applications. The observed limitations in detecting life-saving appliances highlight the importance of dataset balancing and tailored augmentation strategies for small object classes in future work.

4.4 Confusion Matrix and Precision Recall Graphs

A confusion matrix is a two-dimensional representation that illustrates a model's performance by enumerating correct and incorrect predictions across all classes. Figures 11a and 11b present the confusion matrices for YOLOv11n and YOLOv8n, respectively. These matrices enable a more granular evaluation of class-wise prediction accuracy, offering valuable insights into each model's strengths and weaknesses in object classification, particularly for critical categories.

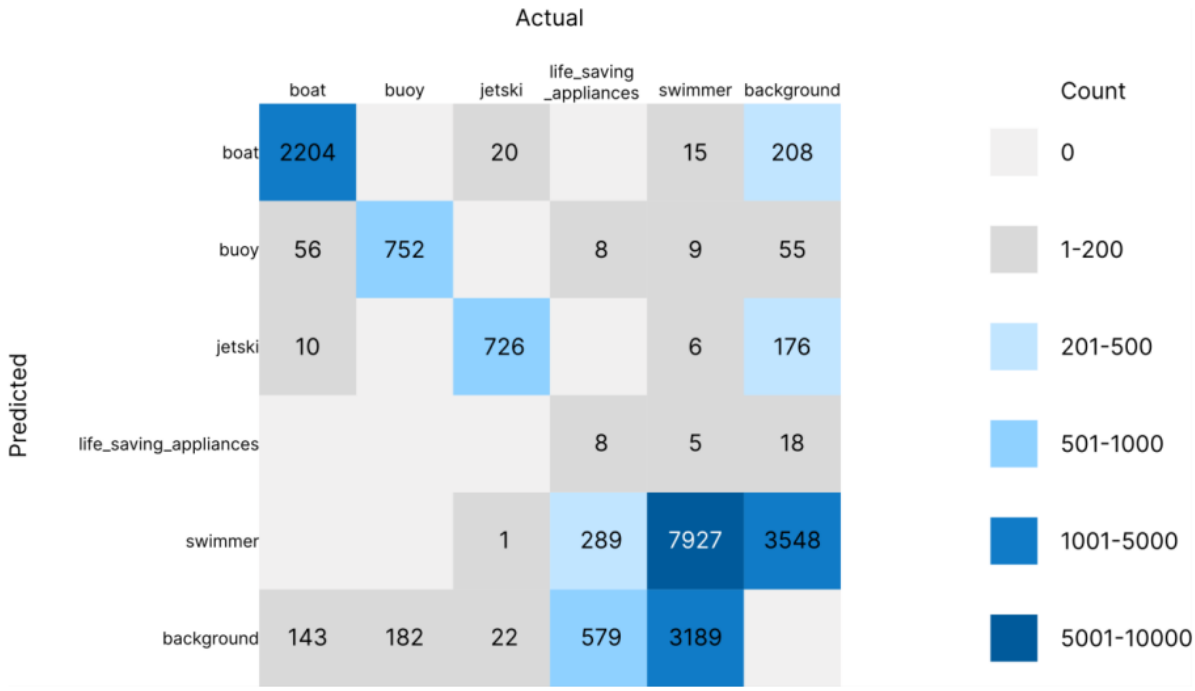

Figure 11a. Confusion Matrix for YOLOv8n

Figure 11b. Confusion Matrix for YOLOv11n

The confusion matrices of YOLOv11n and YOLOv8n reveal consistent patterns in maritime object detection. Both models achieve strong correct classification for larger, visually distinct classes such as boats and jetskis, highlighting their robustness in identifying prominent objects. In contrast, swimmers, the primary focus of this study, exhibit higher confusion with other classes, underscoring the challenge of detecting smaller, dynamic targets in complex maritime environments. The life_saving_appliances class remains the most difficult to classify accurately, with minimal correct identifications, likely due to its small size and low validation set representation.

Interestingly, the distribution of diagonal elements (correct predictions) in the confusion matrices is notably similar between the two models. This suggests that despite architectural differences, YOLOv11n and YOLOv8n approach maritime object detection with comparable performance boundaries and limitations.

The overall accuracy for both YOLOv8n and YOLOv11n models, as derived from their respective confusion matrices, are presented in Table 8.

Table 8. Overall accuracy of both the models

Model	Accuracy
YOLOv8n	0.492
YOLOv11n	0.481

The precision-recall (PR) curves of YOLOv8n and YOLOv11n, depicted in Figures 12a and 12b respectively, illustrate the trade-offs between precision and recall across varying confidence thresholds. Both models exhibit the typical trend seen in object detection tasks, precision decreases as recall increases, highlighting the trade-off between minimizing false positives and maximizing object detection.

At lower recall values, both models maintain high precision, which gradually declines as the models strive to detect more instances. The area under the curve (AUC) is relatively balanced between the two models, indicating comparable overall performance. However, slight differences in the shape of the curves suggest nuanced variations in how each model approaches maritime object detection, particularly in complex classes such as swimmers and life_saving_appliances. These differences emphasize the need to evaluate both precision and recall together when selecting a model for practical, real-time search and rescue scenarios.

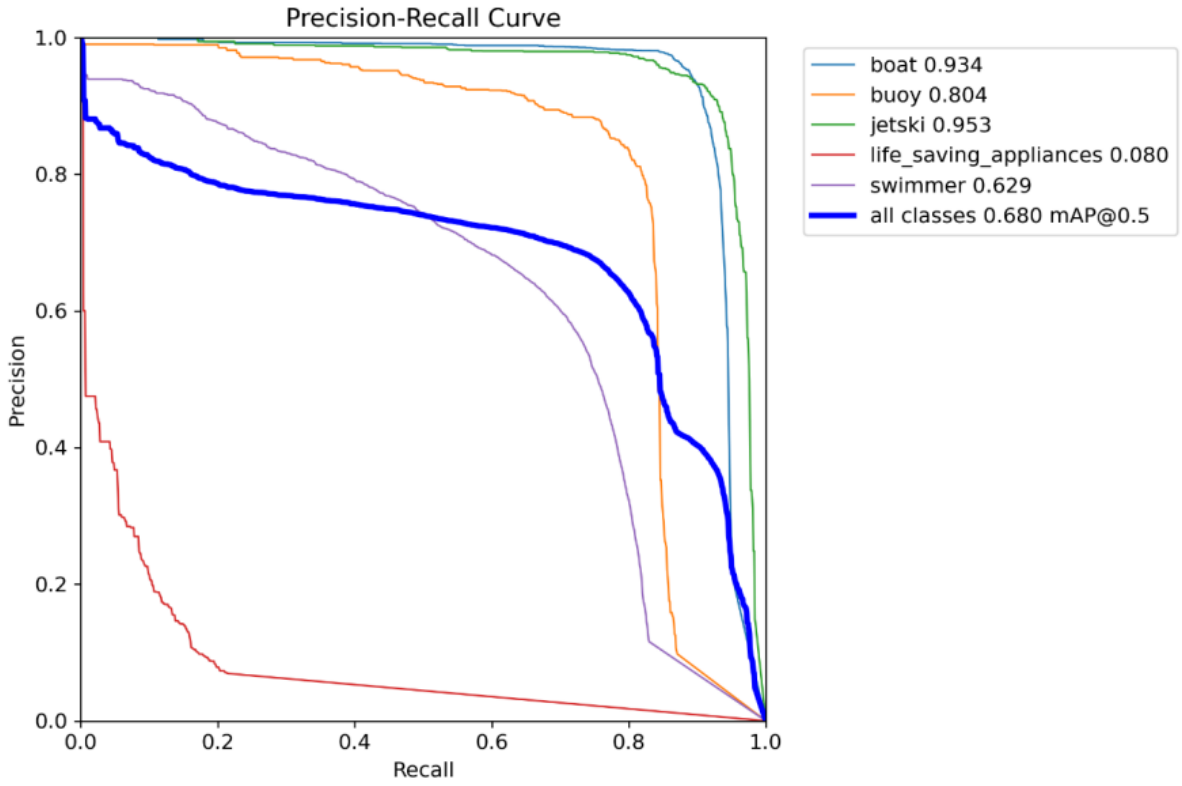

Figure 12a. Precision-Recall curve for YOLOv8n

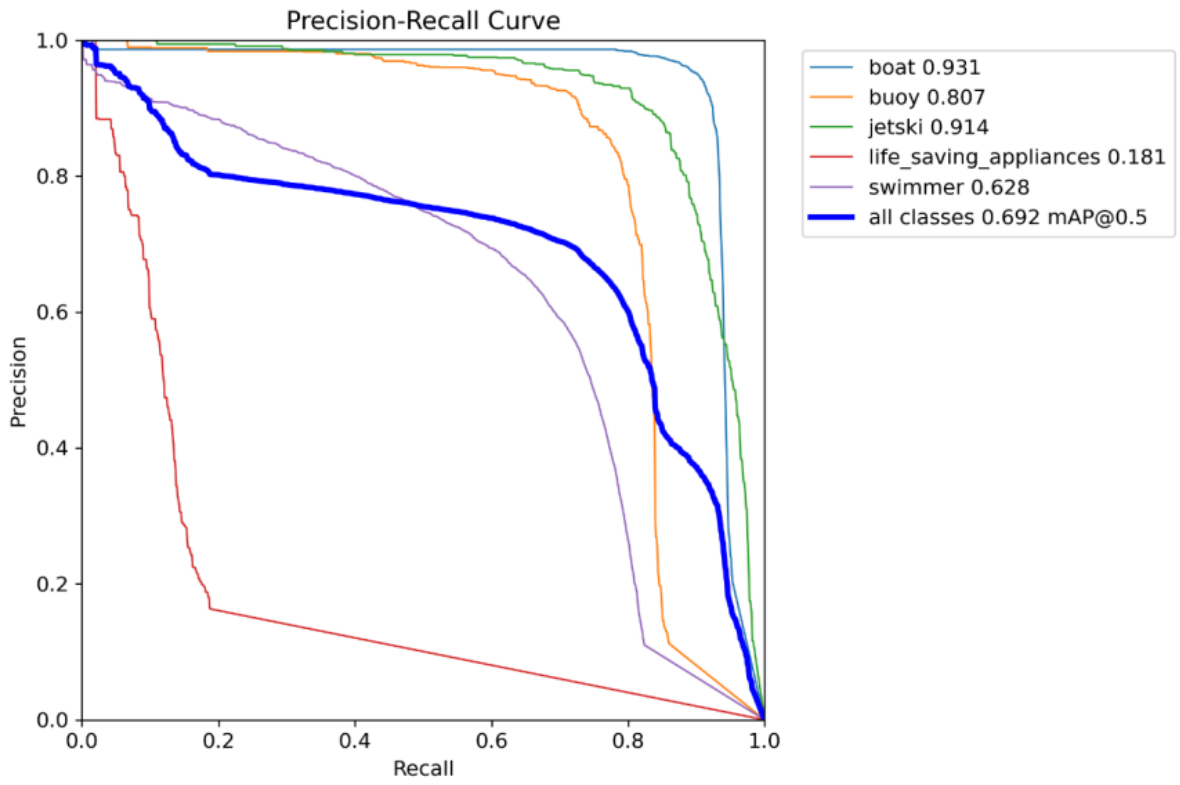

Figure 12a. Precision-Recall curve for YOLOv11n

4.5 Outputs of the Detection Models

Predicted results from the test dataset on the YOLOv8n model are shown in Figures 13a, 13b and 13c.

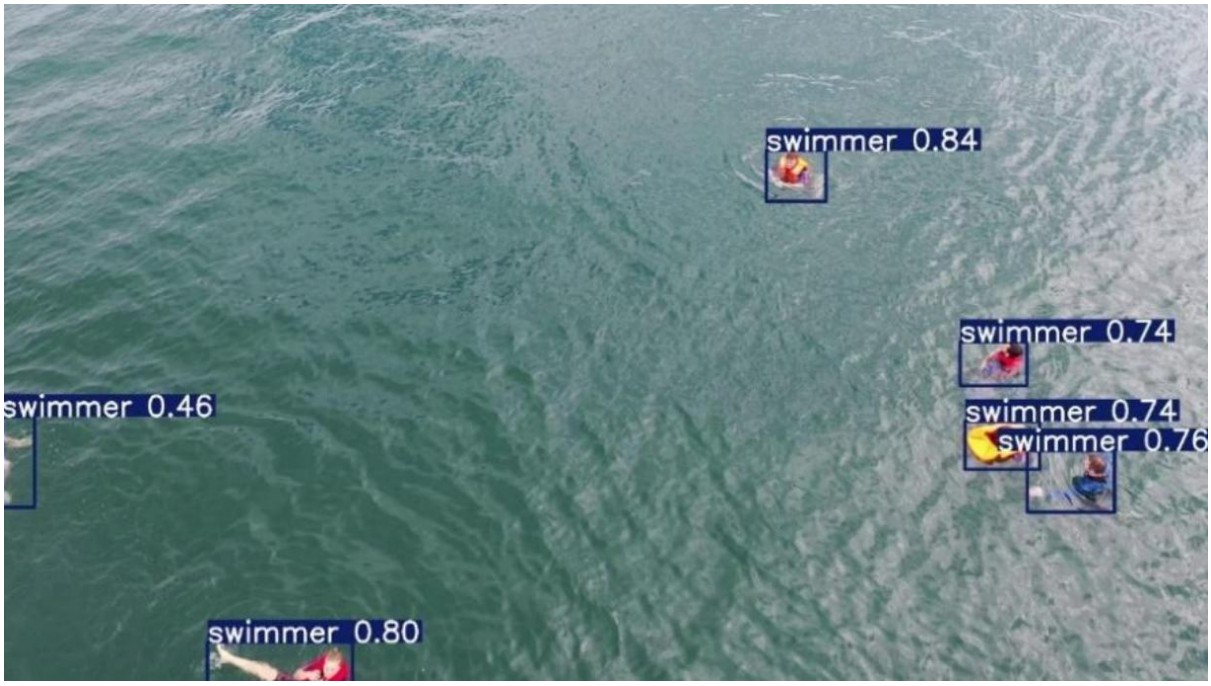

Figure 13a. Detection of Swimmers

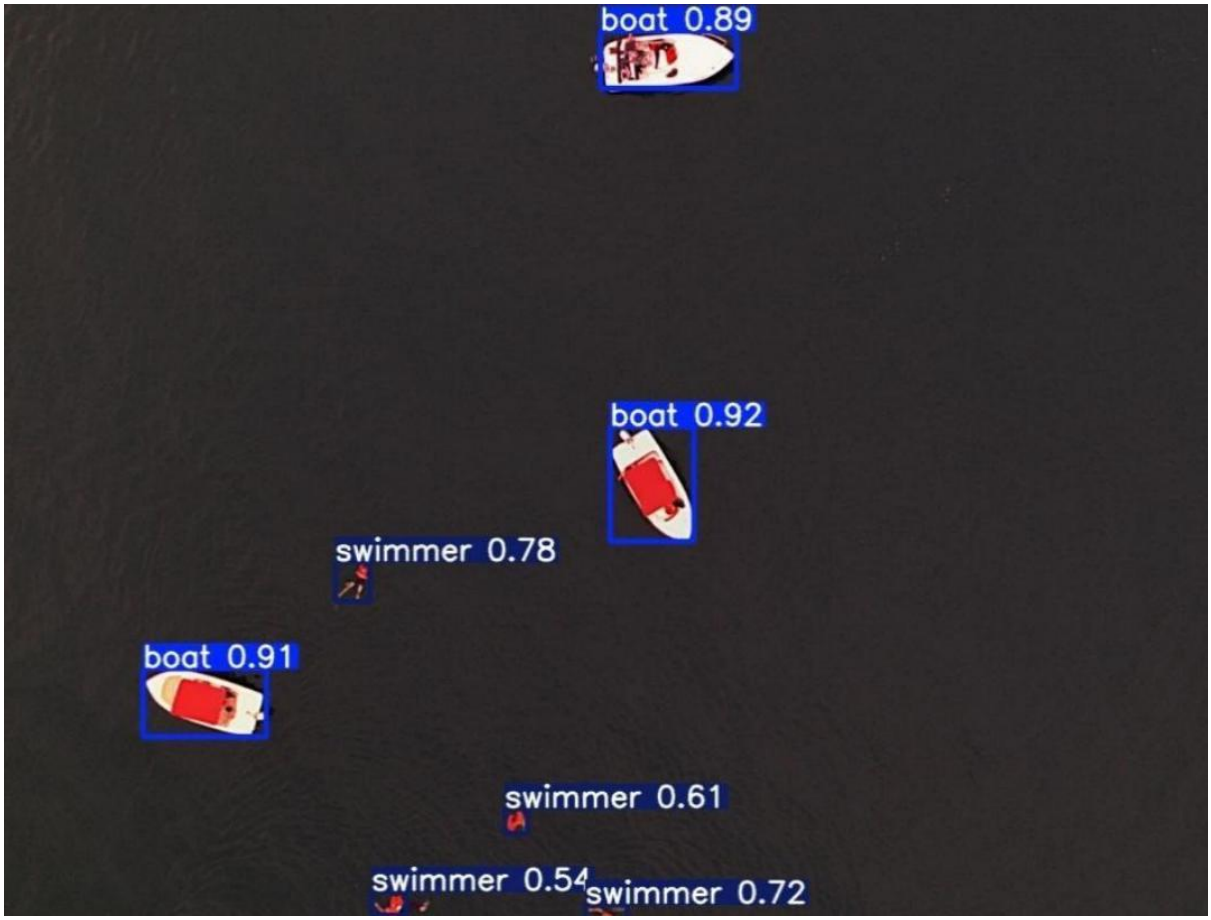

Figure 13b. Detection of Boats and Swimmers

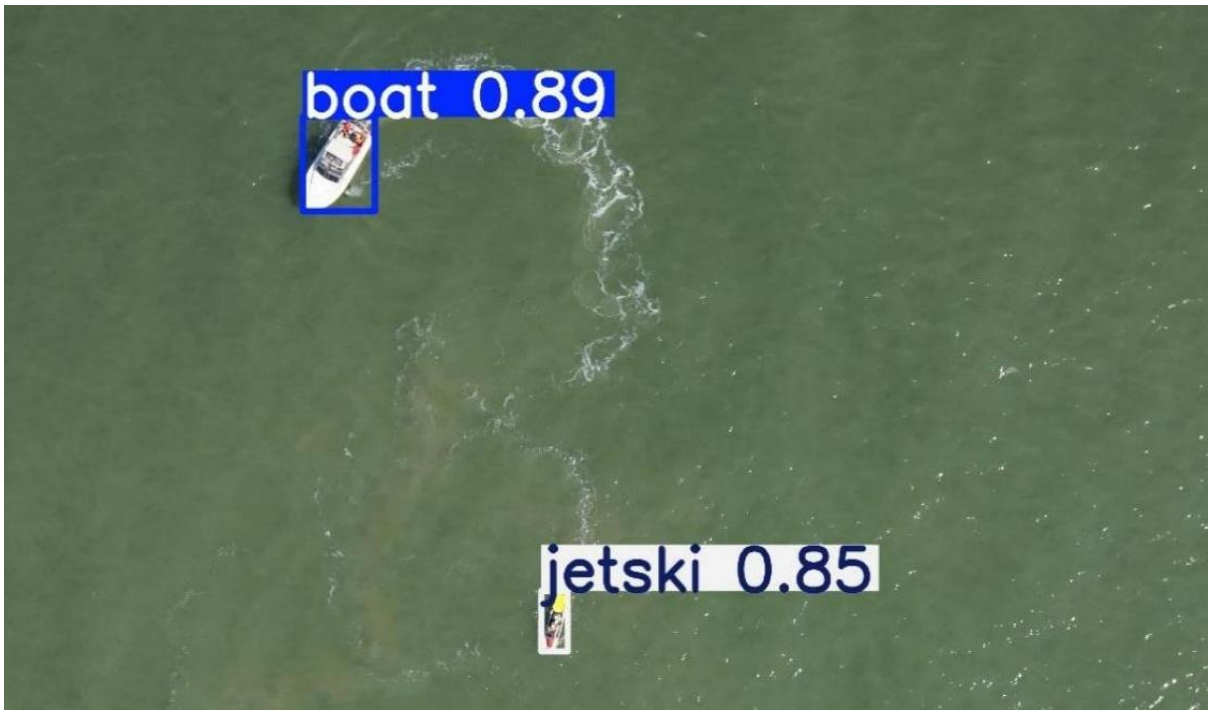

Figure 13c. Detection of Boats and Jetski

The collective comparison of outputs or performances over multiple metrics or visual representations identified as 13a, 13b, 13c, 13d, and 13e reveals variations of results or outcomes under specific conditions, including changes to model parameters, input data, or environments. Analyses of the subfigures reveal a variety of performance differences based on accuracy, precision, and clarity.

Predicted results generated by the YOLOv8n model on real-world maritime scenarios, sourced from publicly available online videos, are presented in Figures 14a, 14b, 14c, 14d and 14e.

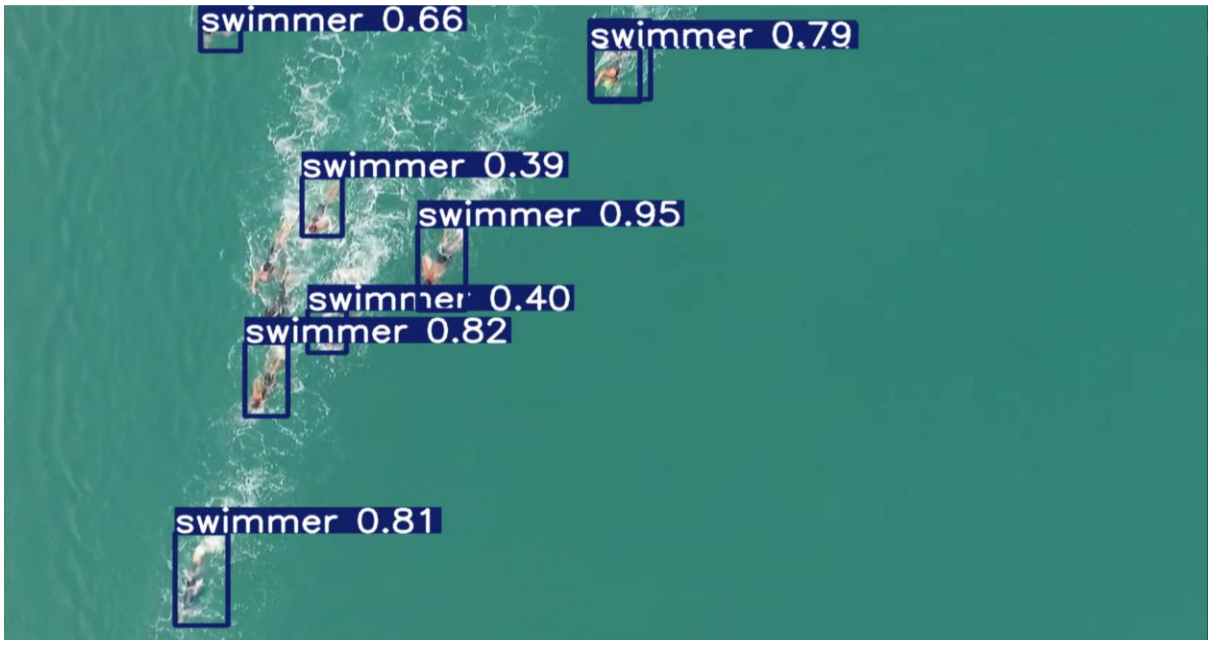

Figure 14a. Detection of multiple Swimmers

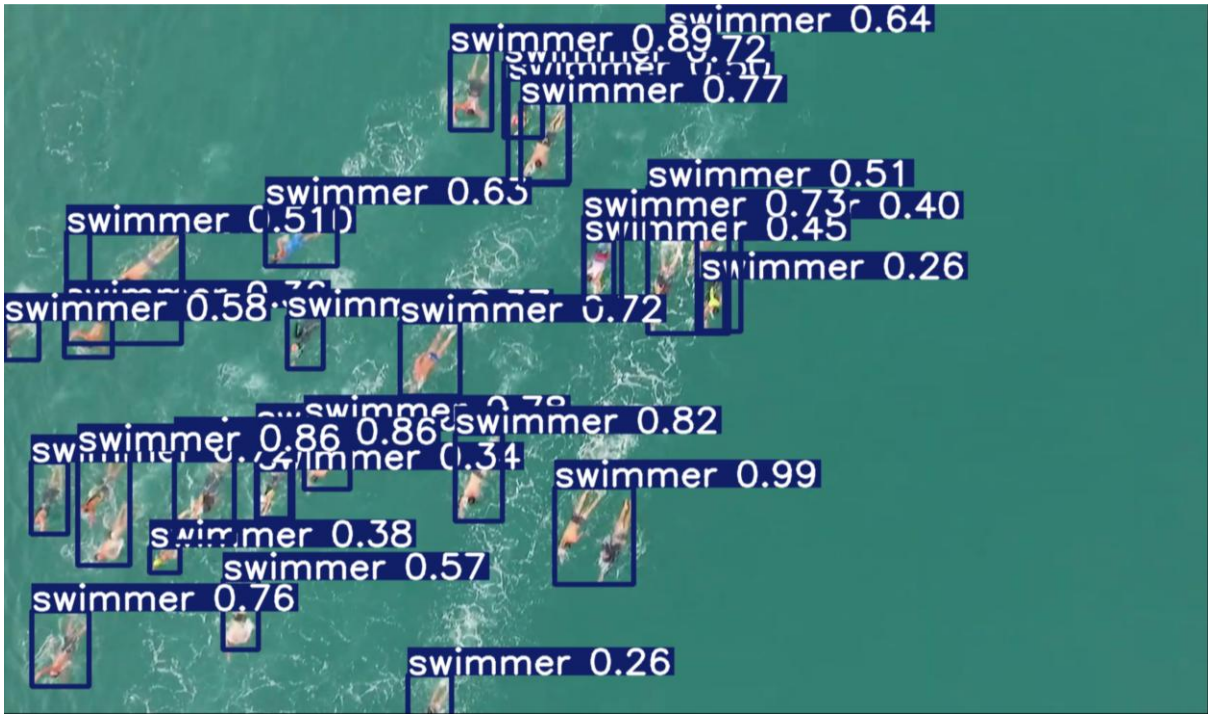

Figure 14b. Detection of multiple Swimmers

Figure 14c. Detection of people on the beach

Figure 14d. Detection of people & boats on the beach

Figure 14e. Detection of Swimmers & Boats in the ocean from Aerial View

Figures 14a–14e present qualitative results of the YOLOv8n model evaluated on randomly sampled video frames from internet sources. These examples illustrate the model's generalization capability in unconstrained, real-world aquatic environments, conditions closely resembling actual marine search and rescue (SARS) scenarios. The qualitative outcomes reveal how the model performs across varying visual complexities, from sparse scenes to crowded marine environments.

In **Figure 14e**, the model is tested on a relatively straightforward case involving a minimal number of swimmers in clear water. YOLOv8n accurately detects and localizes the individual with a tight bounding box and high confidence score. This confirms the model's strong baseline performance under favourable conditions, good visibility, low background noise and minimal occlusion. It also demonstrates the model's precision in detecting human figures when the object-to-background contrast is high.

Figure 14a depicts a more challenging situation involving multiple swimmers in close proximity. The model detects most individuals with correctly scaled bounding boxes and consistent confidence scores. However, one swimmer near the edge of the frame is partially or entirely missed, highlighting a common limitation in edge-region detection. This figure suggests that while YOLOv8n is effective in handling moderate occlusion and density, boundary cases remain a challenge, possibly due to training dataset biases toward cantered objects.

Figure 14b presents a high-density aquatic scene with multiple swimmers in close quarters, some overlapping. Despite significant visual clutter and potential motion blur, the model detects a

majority of the individuals. The bounding boxes in this case are tightly packed, some slightly overlapping, indicating effective spatial localization under crowding. However, a few swimmers in the background remain undetected, likely due to scale reduction and partial occlusion. This indicates that the model may underperform in cases involving small or distant objects, suggesting potential benefits from scale-aware enhancements or multi-resolution features.

Figure 14c and **Figure 14d** collectively illustrate YOLOv8n's effective object detection performance in beachside environments using aerial imagery. In both scenes, the model successfully identifies multiple individuals labelled as "swimmer" with bounding boxes displaying moderate to high confidence levels, ranging from 0.42 to 0.71. The model maintains strong localization accuracy even in regions with groups of people or partial occlusions from natural elements like trees and man-made objects such as vehicles. Notably, Figure 13f also captures a distant "boat" with a reasonable confidence score, showcasing YOLOv8n's versatility across object categories in complex outdoor settings. Across both images, despite some distant figures remaining undetected due to scale, the model demonstrates consistent detection across foreground and mid-ground regions, reinforcing its suitability for real-time monitoring in open, dynamic environments such as coastal or recreational areas.

4.5 Impact of Model Size on Rescue Operations Efficiency

In time-sensitive maritime search and rescue operations, the balance between computational efficiency and detection performance is critical. Lightweight object detection models like YOLOv8n and YOLOv11n are well-suited for deployment on edge devices such as UAVs, USVs, and embedded rescue systems due to their low inference overhead.

YOLOv11n, with a lightweight footprint of just 6.3 GFLOPs, offers exceptional computational efficiency, enabling faster processing speeds and lower energy consumption, key advantages for autonomous marine SAR systems operating on resource-limited platforms like drones. However, this efficiency comes with a slight trade-off in detection performance.

YOLOv8n, although marginally larger at 8.1 GFLOPs, consistently outperforms YOLOv11n across critical metrics including precision (0.818 vs. 0.810), recall (0.601 vs. 0.592), F1 score (0.693 vs. 0.684), mAP_{50} (0.621 vs. 0.611), and accuracy (0.492 vs. 0.481). In high-stakes rescue scenarios where detection accuracy can directly impact outcomes, these improvements are significant.

4.6 Comparison of YOLOv8n and YOLOv11n models

YOLOv11n, with a lower computational load of 6.3 GFLOPs, offers notable efficiency benefits. Its lightweight architecture enables faster inference and lower power consumption, key advantages for

real-time deployment in remote or power-constrained conditions. However, this efficiency comes with a slight compromise in detection performance.

YOLOv8n, while marginally heavier at 8.1 GFLOPs, demonstrates superior detection capabilities across nearly all evaluation metrics. It achieves a precision of 0.818, recall of 0.601, F1 score of 0.693, mAP50 of 0.621, mAP50–95 of 0.393, and accuracy of 0.492. In comparison, YOLOv11n records a precision of 0.810, recall of 0.592, F1 score of 0.684, mAP50 of 0.611, mAP50–95 of 0.369, and accuracy of 0.481. These differences, while moderate, are consistent across key performance indicators and suggest that YOLOv8n is more robust in identifying swimmer instances under varied visual and environmental conditions.

While YOLOv11n remains a strong candidate when prioritizing computational efficiency, YOLOv8n offers a more balanced trade-off between accuracy and model size, making it a more reliable choice for safety-critical deployments where every missed detection can have serious consequences. Therefore, YOLOv8n is the preferred model for deployment in autonomous rescue operations, where both precision and recall are crucial for maximizing swimmer detection coverage.

5. Conclusion

5.1 Summary of Key Findings

The evaluation results indicate that both YOLOv8n and YOLOv11n models deliver competitive performance for the marine search and rescue detection task. YOLOv8n achieves a slightly higher precision of 0.818 compared to 0.810 for YOLOv11n, along with a better recall of 0.601 versus 0.592. This results in an F1 score of 0.693 for YOLOv8n, outperforming YOLOv11n's 0.684, demonstrating a more balanced ability to correctly detect positive cases while minimizing false detections.

In terms of detection accuracy, YOLOv8n also outperforms YOLOv11n with a higher mAP50 of 0.621 versus 0.611, and a better mAP50-95 score (0.393 compared to 0.369), indicating superior precision across various Intersection over Union (IoU) thresholds. Although YOLOv11n benefits from a lower computational cost, measured at 6.3 GFLOPs compared to YOLOv8n's 8.1 GFLOPs, this reduction in complexity comes at the expense of slightly diminished overall detection performance.

Given the critical nature of search and rescue operations where detection accuracy and reliability are paramount, YOLOv8n is preferred despite its marginally higher computational demands. Its improved precision and recall translate into more dependable object detection in real-time scenarios, which is crucial for effective deployment in dynamic marine environments. Therefore, YOLOv8n offers a better overall trade-off, balancing computational efficiency with superior detection performance, making it the more suitable choice for operational use cases requiring high accuracy.

5.2 Limitations of the Study

- The analysis was conducted using specific benchmark datasets that may not fully represent the wide range of conditions encountered during real-world rescue operations, potentially limiting the generalizability of the results.
- The evaluation of computational costs was based on theoretical GFLOPs and did not consider actual deployment on diverse hardware platforms, which could affect the models' performance and efficiency in practical applications.
- The object detection capabilities under challenging environmental conditions such as poor visibility, low light, or adverse weather were not explicitly assessed, leaving uncertainty about model robustness in such scenarios.

5.2 Limitations of the Study

- **Dataset Expansion:** Future studies should focus on developing and evaluating models using more diverse and representative datasets that capture the complexities of real-world rescue operations, including scenarios with poor lighting conditions, adverse weather, and underwater environments.
- **Hardware Optimization:** There is a need to optimize YOLOv11 models for deployment on resource-constrained edge devices, such as drones or embedded systems, to enhance their applicability in field operations with limited computational power.
- **Scenario Testing:** Conducting comprehensive analyses of dynamic and complex environments, such as moving objects, water currents, and floating obstacles, is essential to better understand model robustness and improve detection accuracy in practical rescue situations.
- **Transfer Learning Approach:** Employ transfer learning techniques by pre-training models on related classes like 'life_saving_appliances' and leveraging these pretrained weights to boost detection performance on specific rescue-related object categories.
- **Energy Efficiency:** Future research should investigate the power consumption and overall energy efficiency of these models to enable sustainable and prolonged deployment during extended rescue missions, ensuring operational reliability without frequent recharging or hardware replacement.