https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Oriented Object Detection for Complex Hydrodynamic Features: A Multi-Platform Rip Current Identification System

Albert Català-Gonell[1,2], Jesús Soriano-González[1], Elena Sánchez-García[1], Francisco Fabián Criado-Sudau[1], Josep Oliver-Sansó[1], Valentin Kozlov[3], Khadijeh Alibabaei[3], José Luis Lisani[4], and Àngels Fernández-Mora[5]

[1]Balearic Islands Coastal Observing and Forecasting System (SOCIB), Palma, Spain
[2]SARTI Research Group, Electronic Department, Universitat Politècnica de Catalunya (UPC), Vilanova i la Geltrú, Spain
[3]Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[4]Institute of Applied Computing and Community Code (IAC3), Universitat de les Illes Balears, Palma, Spain
[5]Earth Science Group, Department of Biology, University of the Balearic Islands, Palma, Spain

Corresponding author: Jesús Soriano-González (jsoriano@socib.es)

Abstract. Rip currents are hazardous, fast-moving seaward flows and remain one of the leading causes of rescues and drownings on surf beaches, yet their automated detection remains a significant challenge due to their amorphous, dynamic morphology and the environmental complexity of the surf zone. This study introduces a novel platform-agnostic deep learning–based framework for automated rip current detection from beach imaging platforms, integrating three core contributions: a diverse new dataset, a rigorous architectural benchmark, and a deployable operational tool. We first present RipAID, a comprehensive dataset enriched with multi-platform imagery and multiple viewing angles to ensure scale-invariant learning. Building on this resource, a systematic evaluation of state-of-the-art architectures demonstrates that geometric fidelity is critical; specifically Oriented Bounding Boxes (OBB) significantly outperform standard axis-aligned methods. Our optimized YOLOv11n-OBB model achieves robust performance ($\mathrm{mAP50}$: 0.927), with inference speeds from 2.4 to 60 FPS on hardware ranging from edge devices to GPU workstations. To bridge the gap between research and practice, and ensure that the results are reusable and reproducible, the framework and model weights have been released as an open-source, containerized module (*socib-rip-currents-detection*), providing the coastal safety community with a scalable, ready-to-use and standardized tool for continuous, automated rip current monitoring.

1 Introduction

Rip currents are narrow, fast-moving, jet-like flows of water extending from near the shoreline to beyond the surf zone (MacMahan et al., 2006; Short, 1985). Their formation is driven by nearshore circulation patterns influenced by hydro-morphological forcing factors, including beach morphology, seafloor topography, and wave characteristics. Based on these mechanisms, rip currents are categorized into three main types: hydrodynamically controlled rips, which typically develop on planar, featureless beaches and are transient and unpredictable; bathymetrically controlled rips, which are driven by seabed features and form

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

repeatedly in fixed locations; and boundary-controlled rips, which form against rigid structures and tend to be persistent in both location and duration (Castelle et al., 2016). Depending on these factors, rip current velocities typically range from 0.5 to $1\ \mathrm{ms}^{-1}$, though they can exceed $3\ \mathrm{ms}^{-1}$ under extreme conditions (Short, 2007).

Rip currents constitute the primary focus of beach lifeguard efforts, remain a leading cause of rescues and drownings globally (Surf Life Saving Australia, 2019), and are considered the principal physical hazard to recreational bathers on surf beaches (Brander and Scott, 2018). Despite their prevalence, many beachgoers remain uninformed of the existence of rip currents or are unable to visually identify them (Caldwell et al., 2013). This lack of awareness often leads individuals to enter the water through a rip channel, as the area frequently displays fewer breaking waves and appears calmer than the surrounding surf. While lifeguards are trained to identify these hazards, coverage is often limited to specific seasons, hours, and locations, creating periods of vulnerability, especially outside of peak times or in less-frequented areas. Furthermore, the unpredictable nature and varying flow patterns of rip currents complicate the provision of universal safety advice, highlighting the need for accurate detection and warning systems to mitigate risks and improve beach safety.

Automated detection of rip currents has historically been challenged by their highly dynamic, amorphous nature. From an in-situ observing perspective, various approaches have been proposed, including dye tracking, wave sensors, acoustic velocimeters, current profilers, and drifters (Song et al., 2016; Bowman et al., 1988; Huntley et al., 1988; MacMahan et al., 2005). While effective, field deployment in the surf zone is logistically difficult, expensive, and often restricted in crowded beach environments. Furthermore, these methods require significant technical expertise and are typically applicable only to highly localized instances in time and space (Pitman et al., 2016), limiting them to experimental scenarios. To overcome these constraints, remote sensing offers an efficient and spatiotemporally enhanced alternative. While video imaging has been used to remotely sense the nearshore for over three decades (Lippmann and Holman, 1989), technological advances have exponentially expanded the availability of coastal data. Beyond earth observation satellites, beach imaging technologies have become increasingly diverse and accessible. Fixed beach video-monitoring systems provide systematic, high-frequency intraday imagery with available open-source solutions (Nieto et al., 2010). Aerial monitoring is no longer restricted to traditional plane-based photogrammetry, as affordable Unmanned Aerial Vehicles (UAVs) and sensors are now widely available for coastal surveying (Adade et al., 2019). Public surfcams have become widespread and of great value for beach monitoring (Andriolo et al., 2019), and the ubiquity of smartphones has turned them into powerful field observation tools, capable of providing valuable coastal information (Soriano-González et al., 2024).

The broad availability of coastal imagery has encouraged the development of automated algorithms for detecting rip currents from these data streams. Early approaches relied on traditional image-processing techniques that exploited the characteristic optical signatures of rip currents. For instance, Holman et al. (2006) leveraged the distinct dark, low-intensity gaps that appear within the bright foam of breaking waves in time-averaged exposures (Timex). Building on these optical properties, Pitman et al. (2016) applied pixel-intensity filtering for background subtraction, comparing each frame to a reference parent image to identify anomalies. Liu and Wu (2019) introduced a detection method based on color segmentation, using thresholds on the Hue Saturation Value (HSV) color space to isolate rip-related sediment plumes. However, these rule-based techniques typically depend on fixed thresholds and idealized assumptions, making them highly sensitive to environmental conditions such as sun

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

glint, shadows, or variable water color. Beyond static imagery, video-based techniques have also been explored, demonstrating potential to capture more complex patterns, such as those of flash rips (Dérian and Almar, 2017; Mori et al., 2022). However, these methods require high-quality, temporally stable video sequences, rendering them less robust to environmental variability and more difficult to deploy across diverse hardware systems.

To address the challenges of high spatiotemporal variability in both image conditions and rip currents morphology, research has pivoted towards deep learning-based computer vision. Most studies rely on Convolutional Neural Network (CNN)–based object detection methods. These methods excel at extracting the complex, non-linear patterns essential for localizing rips in variable environments. Deep learning detection methods are commonly categorized into two main families: two-stage detectors and one-stage detectors. Two-stage detectors, such as Faster Region-based CNN (Faster R-CNN)(Ren et al., 2017), process the image in two steps: first, a Region Proposal Network (RPN) generates a set of candidate object regions, and then a separate head performs classification and bounding-box regression for each proposal. In contrast, one-stage detectors, such as You Only Look Once (YOLO) (Redmon et al., 2016), directly predict bounding boxes and class probabilities in a dense, fully convolutional manner, often leveraging multiple feature maps at different scales. This architectural design enables substantially faster inference, making one-stage models the preferred choice for most real-time applications.

In recent years, the application of deep learning to coastal imagery has yielded promising results in rip current detection, yet significant operational limitations persist. Initial implementations, such as de Silva et al. (2021), successfully utilized Faster R-CNN to detect rip currents in Google Earth imagery. Subsequent work by Zhu et al. (2022) addressed the inference speed limitations of this approach by adopting a YOLOv5 architecture. However, both models were trained exclusively on top-down satellite-derived data, limiting their generalization to the oblique, shore-based views typical of standard coastal monitoring systems. Rampal et al. (2022) sought to enhance model interpretability and generalization using Class Activation Maps (CAMs); however, their approach relied exclusively on aerial imagery, and struggled to resolve multiple rip current instances within a single frame. Following this, Rashid et al. (2023) achieved improved detection accuracy of 98.55% by applying an enhanced Tiny-YOLOv3 to time-averaged imagery. Yet, this method's dependency on pre-processed inputs restricts real-time flexibility, as standard video feeds do not natively output Timex products. Most recently, research has focused on increasing precision through advanced modalities. Dumitriu et al. (2025) obtained highly precise form depictions of rip currents using instance segmentation, yet required detailed pixel-level annotations, difficult to scale for amorphous and highly variable targets. Similarly, Khan et al. (2025b) achieved high accuracy (94.8%) using active drone imagery, but the necessity for active deployment and favorable flight conditions renders the system unsuitable for continuous, unattended coastal monitoring.

The objective of this study is to contribute to bridging these gaps by developing a robust rip current detection application capable of operating on standard beach imaging platforms, including smartphones, beach videomonitoring systems, and UAV imagery. This approach aims to provide a scalable, cost-effective alternative for beaches without lifeguard coverage, or a complementary tool for supervised areas, enhancing safety through continuous, automated rip current monitoring. To this end, this study introduces a platform-agnostic framework for automated rip current detection, grounded in three core contributions:

- The curation of RipAID v2.0.0, a comprehensive multi-source training dataset.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

- An architectural benchmark comparing state-of-the-art detectors and identifying the role of annotation geometry.
- The development of a ready-to-deploy inference module optimized for real-time operational framework.

The remainder of this paper is organized as follows: Section 2 details the dataset composition, deep learning architectures, and experimental design. Section 3 presents the results, ranging from baseline model comparisons to hyperparameter optimization and cross-validation. Section 4 discusses the implications of these findings for operational monitoring, highlighting the advantages of the proposed approach while identifying directions for further research. Finally, Section 5 summarizes the main conclusions of this work.

2 Methods

2.1 RipAID training dataset

The '*Rip current Annotated Image Dataset*' (RipAID v2.0.0; (Soriano-González et al., 2026)) was constructed from three distinct data sources (Soriano-González et al., 2025; CnE UFSC , 2023; Khan et al., 2025b), and includes 6,789 images and a total of 10,131 annotations. The images were captured across various beaches using multiple image acquisition systems (*e.g.,* smartphones, orthophotos, fixed stations), encompassing a wide range of conditions, including different angles, variable heights, orientations, fields of view, rip current characteristics, and diverse meteoceanic and lighting conditions. Only instantaneous RGB images (*i.e.,* snapshots) are included in RipAID, ensuring alignment with the capabilities of most camera systems (*e.g.,* surfcams, Argus). RipAID annotations are limited to rip currents identifiable in images, that are those appearing as calm channels between breaking waves or as sediment plumes extending seaward.

The dataset labels comprise three classes, *rip current*, *doubt*, and *sediment*, annotated with Oriented Bounding Boxes (OBBs). These are similar to standard bounding boxes (*i.e.,* axis-aligned rectangles that enclose the object to be detected), but rotated to better fit the shape and orientation of the object of interest (see Fig. 1). RipAID v2.0.0 dataset features 4013 *rip current* instances (present in 53% of images), alongside 4,591 *sediment* (23%) and 1,437 *doubt* (19%) instances, with 16% of images lacking any annotations. The label *rip current* indicates clearly visible and identifiable rip currents; *doubt* denotes features that exhibit some rip current characteristics, but the annotator could not confirm them with certainty (*i.e.,* not clearly identifiable); and *sediment* refers to sediment plumes that might relate to a rip current. For further details, readers are referred to the RipAID v2.0.0 Readme file (Soriano-González et al., 2026).

2.2 Deep learning methods

To detect rip currents, one-stage and two-stage detectors were evaluated. This section details the deep learning models used, their type and main distinguishing features.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

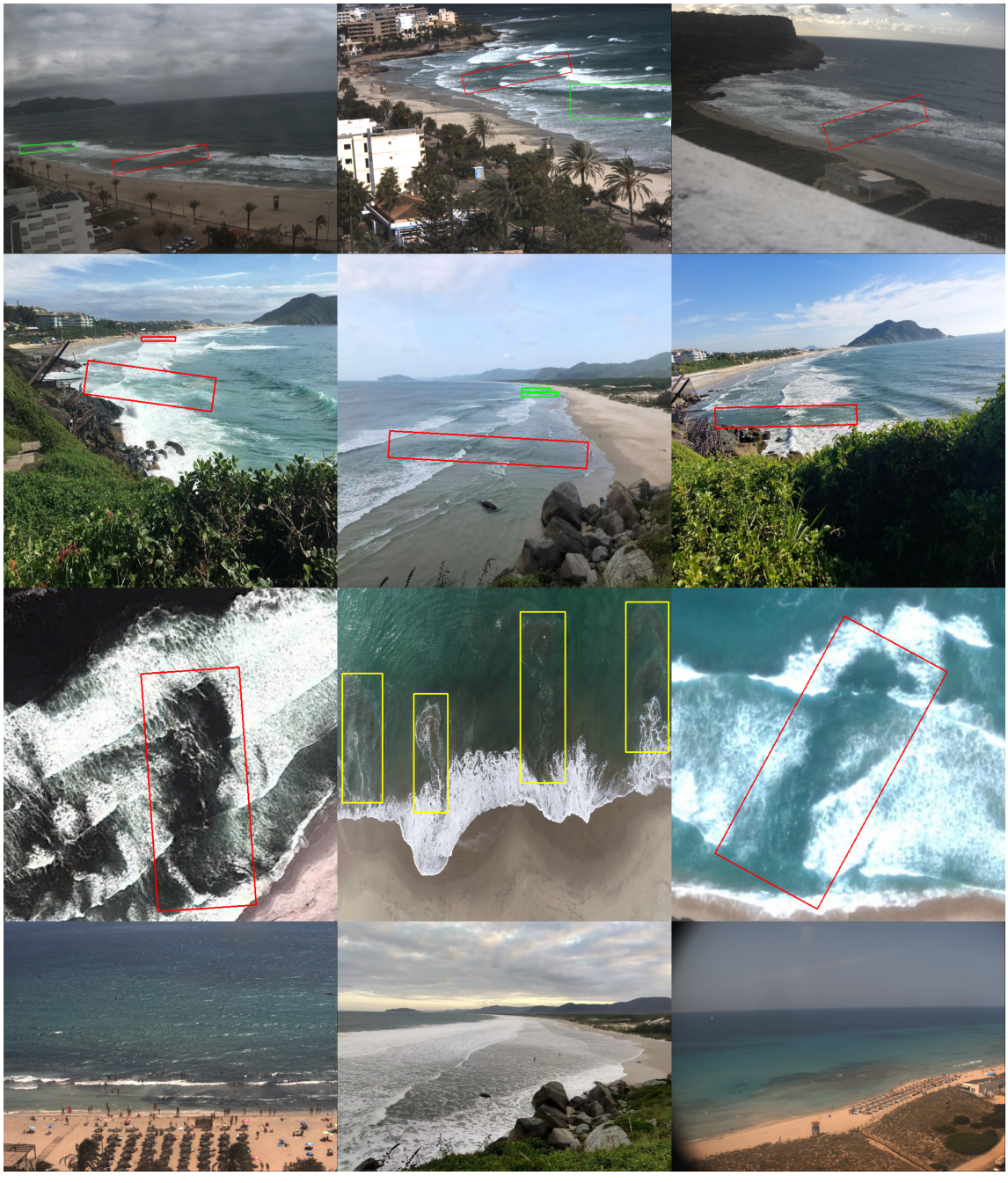

Figure 1. Examples of annotated images from the training dataset, showing rip currents (red), doubt (green), and sediment (yellow). The first three rows contain images with rip currents, while the last row shows images without any rip current annotations.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

2.2.1 Faster R-CNN - Faster Region-based Convolutional Neural Network

Faster R-CNN (Ren et al., 2017) is a fundamental two-stage detector. It employs a Region Proposal Network (RPN) to generate candidate regions, which are subsequently refined by a Fast R-CNN head for classification and bounding box regression. Our implementation incorporates Feature Pyramid Networks (FPN) to enhance multi-scale detection and adopts the architectural enhancements proposed by Li et al. (2021), including dual RPN layers and batch normalization. This configuration results in a model size of approximately 43.7M trainable parameters.

2.2.2 YOLO - You Only Look Once

The YOLO family (Redmon et al., 2016) represents the state-of-the-art in real-time, one-stage object detection across various research domains such as medicine, agriculture, and industry (e.g., Khan et al., 2024; Jayagopal et al., 2024; Lv et al., 2023) and has been also previously used in the framework of rip current detection (Zhu et al., 2022; Rashid et al., 2023; Dumitriu et al., 2023). In this study, the recent OBB supported iterations,YOLOv8 (Jocher et al., 2023a), YOLOv11 (Jocher and Qiu, 2024), and YOLOv12 (Tian et al., 2025), were evaluated across varying model scales ranging from Nano (`n`) to Extra-Large (`x`), with the smallest variants containing as few as 3.5M parameters. The support of OBB detection allows the models to capture directional components of objects. Implemented via the Ultralytics library (Jocher et al., 2023b), these models predict bounding boxes and class probabilities in a single dense pass.

2.2.3 FCOS - Fully Convolutional One-Stage object detector

FCOS (Tian et al., 2019) is an anchor-free model that directly predicts bounding boxes at each pixel location on the feature map, eliminating region proposals and anchor boxes. It treats each location in the feature map as a potential object center, estimating distances to box boundaries and class probabilities. Like Faster R-CNN, FCOS uses FPN for multi-scale detection, assigning small objects to high-resolution layers and large objects to low-resolution layers. This approach simplifies the architecture (32.3M trainable parameters) and reduces computational cost. It also includes a center-ness module, which suppresses low-confidence predictions that are far away from object centers. However, it still needs post processing techniques like Non-Maximum Suppression (NMS) for clean outputs.

2.2.4 RT-DETR - Real-Time Detection Transformer

RT-DETR (Zhao et al., 2024) is the first real-time, end-to-end object detector based on the Transformer architecture introduced in Carion et al. (2020). It replaces the conventional convolutional detection head with a Transformer decoder, enabling the modeling of global context and direct prediction of object sets. A defining advantage of this architecture is the elimination of NMS as a post-processing step, which streamlines the inference pipeline and reduces latency. We utilized the Ultralytics implementation, evaluating the Large (RT-DETR-L, 42M parameters) and Extra-Large (RT-DETR-X, 74M parameters) variants (Jocher et al., 2023b).

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

2.2.5 RetinaNet

RetinaNet (Lin et al., 2020) is a single-stage object detection model designed to address the accuracy limitations of earlier one-stage detectors while maintaining high inference speed. Its main contribution is the Focal Loss, a novel loss function that reduces the impact of easily classified background examples during training, allowing the model to focus on hard examples and improving performance on rare or small objects. Like Faster R-CNN and FCOS, it uses a FPN backbone for multi-scale detection. The version used in this work incorporates improvements from Zhang et al. (2020), including Adaptive Training Sample Selection, which automatically defines positive and negative samples based on object statistics, enhancing accuracy without adding significant computational overhead. This implementation contains 38M parameters and balances speed with robust performance across objects of varying scales and occlusion levels.

2.3 Metrics

Models performance was evaluated using common detection metrics (precision, recall, mean Average Precision), and the self-defined metric $\mathrm{rip\_fit}$. All are defined using fundamental detection concepts (Zhu et al., 2020; Padilla et al., 2021), and are described in Table 1.

Table 1. Definition of the detection metrics used in this study

Metric	Definition
Intersection over Union (IoU)	Calculated by dividing the area of overlap between the predicted bounding box and the annotated one by the area of their union. This gives a value between 0 and 1, and a detection is considered correct if it has an IoU greater than a selected threshold (the most common one is 0.5).
True Positives (TP)	Correctly identified detections.
False Positives (FP)	Detections made in the background.
True Negatives (TN)	Correctly excluded detections.
False Negatives (FN)	Objects that were not detected.
Precision	Indicates how many of the total detections are correct. $\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$.
Recall (or sensitivity)	Indicates how many of the total rip currents are identified. $\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$.
mean Average Precision (mAP)	AP: Area under the curve defined by precision as a function of recall at the selected IoU threshold. mAP: mean AP across all classes. Often noted with the IoU threshold(s) used (*e.g.,* $\mathrm{mAP50}$, $\mathrm{mAP50-95}$).
$\mathrm{rip\_fit}$	Metric introduced to emphasize recall, essential for safety-critical applications, while still accounting for precision to maintain an appropriate precision–recall trade-off. Defined as: $\mathrm{rip\_fit} = 0.5 \times \mathrm{mAP50} + 0.5 \times \mathrm{Recall}$.

2.4 Experiments

To identify the optimal detection framework, we followed a hierarchical experimental approach consisting of three distinct phases: (1) a baseline architectural comparison, (2) an evaluation of annotation modalities (axis-aligned vs. OBB), and (3) an

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

assessment of class-configuration sensitivity (Permissive vs. Conservative). This stepwise approach served as a filter, narrowing the pool of candidate models based on empirical performance before proceeding to subsequent, more granular analyses. Across all phases, models were evaluated using the same training dataset and consistent data partitions (70% training, 20% validation, 10% test) to ensure comparability. For the initial experiments (Sections 2.4.1 and 2.4.2), we adopted a “permissive" annotation strategy, treating ambiguous instances (classified as “doubt" or “sediment") as positive rip current samples.

2.4.1 Baseline- all vs all with axis-aligned annotations

This experiment established a performance baseline by evaluating all candidate architectures under identical experimental conditions. To ensure broad compatibility across all model types including one-stage, two-stage, and transformer-based, standard axis-aligned bounding boxes were used. A total of twenty model variants were trained, including Faster R-CNN, FCOS, RetinaNet, and all scaled variants (Nano through Extra-Large) of RT-DETR, YOLOv8, YOLOv11, and YOLOv12. All models were initialized with pre-trained weights and trained for 150 epochs with a batch size of 16 using default training hyperparameters.

2.4.2 Annotation type- Standard vs OBB

This experiment evaluated the impact of OBB annotations on detection performance, with specific focus on rip currents where the aspect ratio and orientation are highly variable. We selected only model families compatible with OBB (YOLOv8, YOLOv11, and YOLOv12) and compared their performance against their axis-aligned variants from Phase 1. Training conditions remained consistent (150 epochs, batch size 16). While YOLOv8-OBB and YOLOv11-OBB were initialized with pre-trained weights, the YOLOv12-OBB variants were trained from scratch, as pre-trained OBB implementations were not publicly available at the time of experimentation.

2.4.3 Annotation class configuration- Permissive vs Conservative

The final experiment assessed the sensitivity of the model to label ambiguity by contrasting two distinct annotation strategies: (i) a conservative configuration, where only clearly identifiable rip currents were treated as positive samples (“doubt” and “sediment” were excluded and considered as background); and (ii) a permissive configuration, where both rip currents and ambiguous instances ("doubt" and "sediment") were categorized positive rip current samples (*i.e.,* Consistent with experiment 2.4.1 and 2.4.2). The models used in this experiment were selected based on their superior performance metrics from the previous two experiments.

2.4.4 Hyperparameter optimization

Hyperparameter tuning aims to determine the parameter values that maximize a selected performance metric for a given model. In this study, tuning was performed to maximize the $\mathrm{rip\_fit}$ self-defined metric (Table 1) using Optuna, an open-source hyperparameter optimization framework (Akiba et al., 2019). Optuna uses a smart sampling strategy, analyzing past trials to early-discard unpromising ones during training (*i.e.,* pruning), significantly reducing the overall optimization time

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

while focusing resources on the most promising configurations. The optimization process involved training parameters such as weight decay and distributed focal loss, and data augmentation parameters related to colour-space transformations, geometric adjustments, and image orientation. These are detailed in the appendix (Table A1). The optimization process comprised 100 trials (*i.e.,* training runs with a specific set of hyperparameters), with each one trained for a maximum of 80 epochs. Initial exploration included 20 start-up trials exempt from pruning to establish baseline performance. For the remaining trials, Optuna uses a statistical method (median pruning strategy), which compares the intermediate results of the current trial to the median of previous completed trials. To avoid wasting resources on poor configurations, trials that perform significantly below the median after a predetermined minimum number of epochs (ten epochs in this study) are pruned.

2.4.5 CrossValidation

In order to get a more reliable estimate of the performance of the model and stability, a 10-fold cross-validation procedure using the optimized model configuration was implemented. The entire dataset was partitioned into ten subsets (*i.e.,* folds) and the model was trained and evaluated ten times, with each iteration using a different set of folds for training (70%), validation (20%), and testing (10%). While the final fold contained a slightly different number of test images due to the total dataset size, the distribution of object instances across the train, validation, and test sets closely mirrored the target 70/20/10 ratio (Fig. 2). This ensured that each run was executed using a balanced and representative data sample.

2.4.6 Workflow

This research was enabled by the iMagine platform[1]. Built upon the software stack developed in the Artificial Intelligence for the European Open Science Cloud (AI4EOSC) project(Heredia et al., 2025), the iMagine platform provides a comprehensive suite of services and tools for image-based analysis, covering the entire machine learning lifecycle, including model development, training, deployment, virtual machine access, and specialized support tools (Azmi et al., 2025b). For dataset preparation, an iMagine cloud instance of the Computer Vision Annotation Tool (CVAT.ai Corporation, 2023) was used. Experiment development was carried out in a virtual environment configured with five CPU cores (10,000 MHz base frequency), 16 GB of system memory, and an NVIDIA Tesla V100-PCIE-32GB GPU. Experiment progress was monitored using a cloud-instance of MLflow, an open-source platform for managing and tracking the machine learning workflow (Zaharia et al., 2018). The final model and module are hosted by the AI4OS-hub organization[2] and are also provided as a Docker image on the iMagine platform, enabling users to perform inference directly. A depiction of the overall workflow is shown in Figure 3.

[1]https://www.imagine-ai.eu/services/imagine-ai-platform
[2]https://github.com/ai4os-hub

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

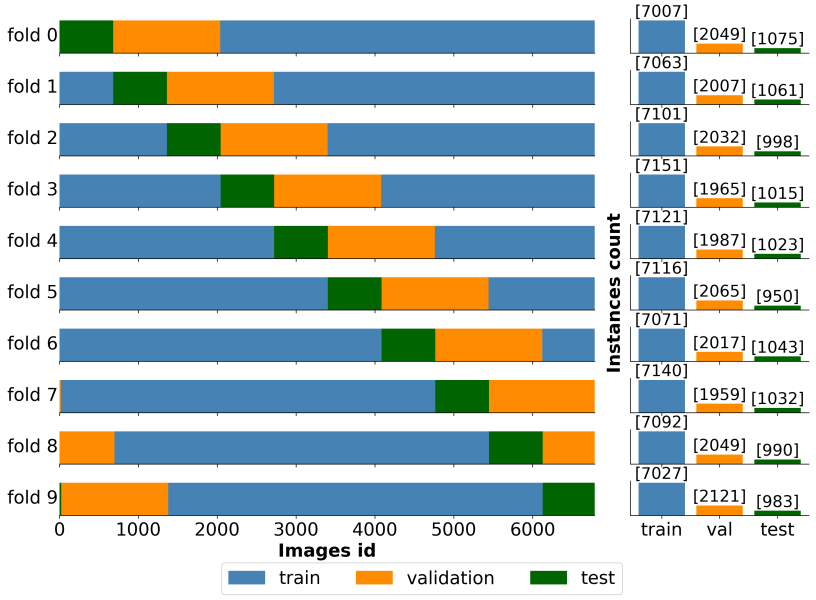

Figure 2. Cross-validation data-folds distribution. Left: Image-split partitioning across folds. Right: Distribution of annotated instances within each split for each fold. Fold 0 corresponds to the baseline dataset distribution used in all experiments previous to crossvalidation.

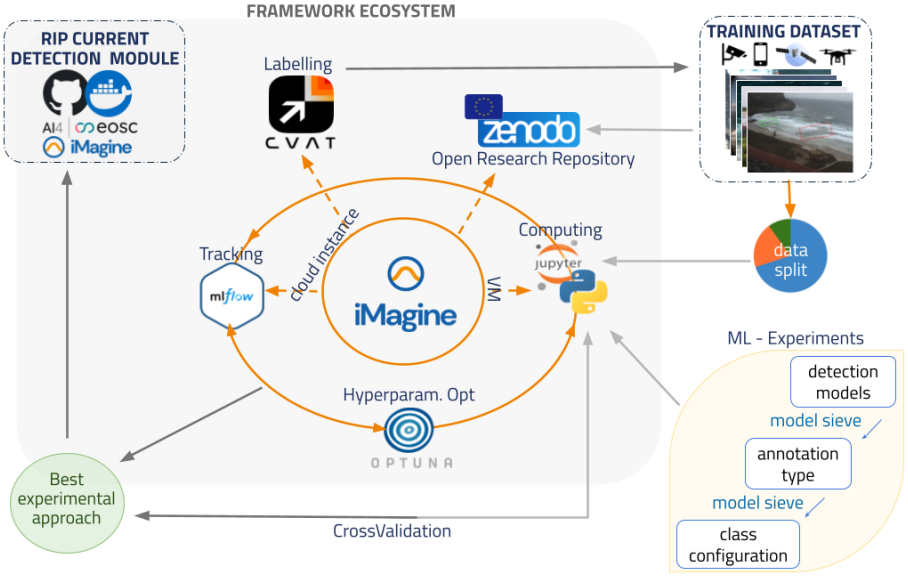

Figure 3. Schematic workflow of the framework ecosystem and the experimental approach, including data preparation, deep learning benchmarking, and inference-module deployment via the AI4EOSC open-access hub.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

3 Results

This section presents the results obtained across all experiments, the hyperparameter optimization, the cross-validation, and the final, packaged module for rip current detection. The evaluation focuses on $\mathrm{mAP50}$, $\mathrm{mAP50-95}$, precision, recall, and the $\mathrm{rip\_fit}$ metrics obtained with the test split.

3.1 Baseline experiment

The baseline experiment assessed all axis-aligned bounding box deep learning models, treating both "doubt" and "sediment" annotations as rip currents. Each model version and size was evaluated individually and their performances can be seen in Table A2. Figure 4 compares the performance of the different considered models across the performance metrics. Architectures evaluated across multiple sizes are represented as violin plots to visualize performance distribution, while single-size models are depicted as individual points. Results show that YOLO and RT-DETR outperform FCOS, Faster R-CNN, and RetinaNet across all metrics except recall, where all models perform comparably (Fig. 4 and Table A2). YOLO models generally achieved slightly higher scores than RT-DETR, except for $\mathrm{mAP50-95}$, where YOLO outperformed RT-DETR margin of 9.3%. The best-performing models in each metric were: YOLOv12l for $\mathrm{mAP50}$ (0.860) and $\mathrm{rip\_fit}$ (0.823), YOLOv11x for $\mathrm{mAP50-95}$ (0.477), YOLOv8x for precision (0.864), and Faster R-CNN for recall (0.839), although the latter incurred with a considerable trade-off in precision (0.545).

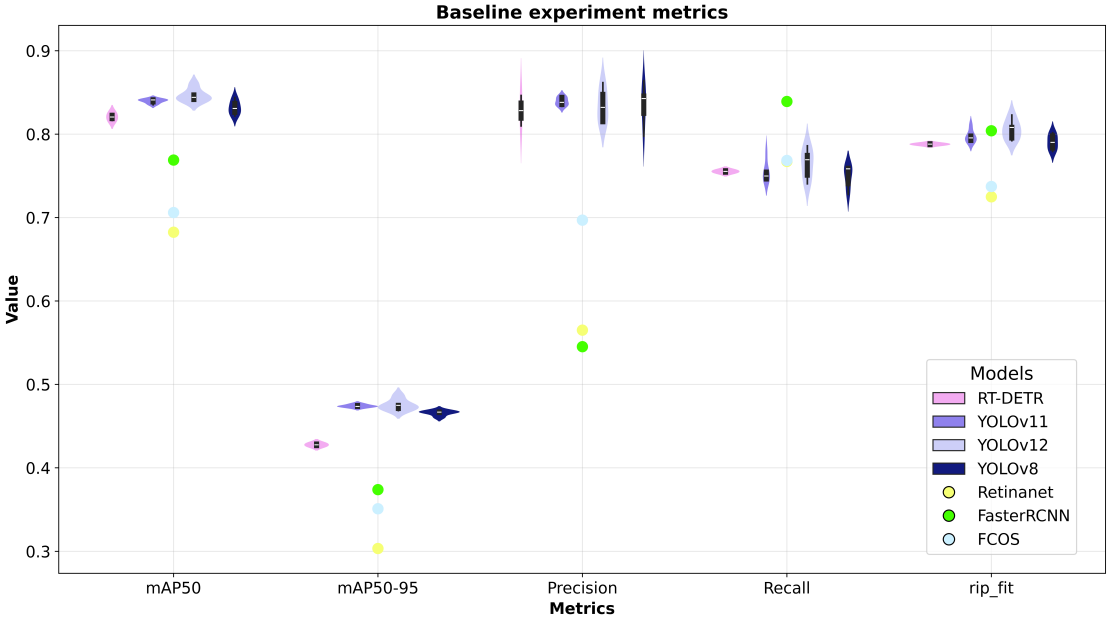

Figure 4. Comparative performance across $\mathrm{mAP50}$, $\mathrm{mAP50-95}$, precision, recall and $\mathrm{rip\_fit}$ metrics in the baseline experiment. Note: The FCOS recall values overlap significantly with those of RetinaNet.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

3.2 Annotation type experiment

The annotation type experiment evaluated OBB models under the same conditions as the baseline experiment. This approach enabled both the identification of the best-performing OBB models and a direct comparison between axis-aligned and oriented annotations in the context of rip current detection. This experiment was restricted to the YOLO family as these models were the only architectures evaluated under both configurations and achieved the highest accuracy metrics in the prior experiment.

Similarly to the baseline experiments, all the models were evaluated individually and they all achieved comparable metrics (see Fig. 5 and Table A3). The models with best results in each metric were: YOLOv8n-OBB for $\mathrm{mAP50}$ (0.916), YOLOv11l-OBB for $\mathrm{mAP50-95}$ (0.652), YOLOv8s-OBB for precision (0.904), and YOLOv11n-OBB for recall (0.843) and $\mathrm{rip\_fit}$ (0.879).

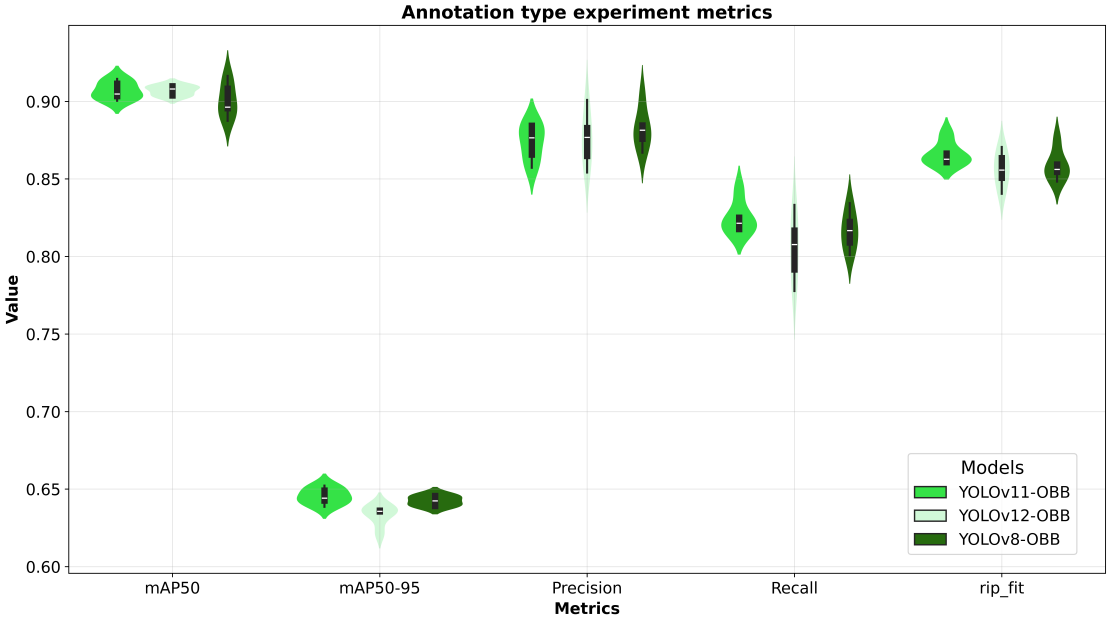

Figure 5. Violin plot showing the performance of each model group in the second experiment across the metrics $\mathrm{mAP50}$, $\mathrm{mAP50-95}$, precision, recall and $\mathrm{rip\_fit}$.

In the comparison of annotation configurations, YOLO-OBB models consistently outperformed their axis-aligned counterparts. YOLO-OBB configurations yielded average performance improvements of 7.8% in $\mathrm{mAP50}$, 35.7% in $\mathrm{mAP50-95}$, 4.9% in precision, 7.9% in recall, and 7.8% in $\mathrm{rip\_fit}$.

As larger models are generally slower and less suitable for near real-time applications, an in-depth analysis was made in order to compare the performance of each model as a function of its parameter count. It compared the $\mathrm{rip\_fit}$ values of all YOLO model sizes and versions with their number of parameters (Fig. 6). The results showed that increasing model size (*i.e.,* parameter count) does not necessarily translate into improved detection performance, and emphasized the clear performance

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

gap between OBB and axis-aligned models. For instance, the YOLOv8 Nano version outperformed the YOLOv8 X-Large in terms of rip_fit.

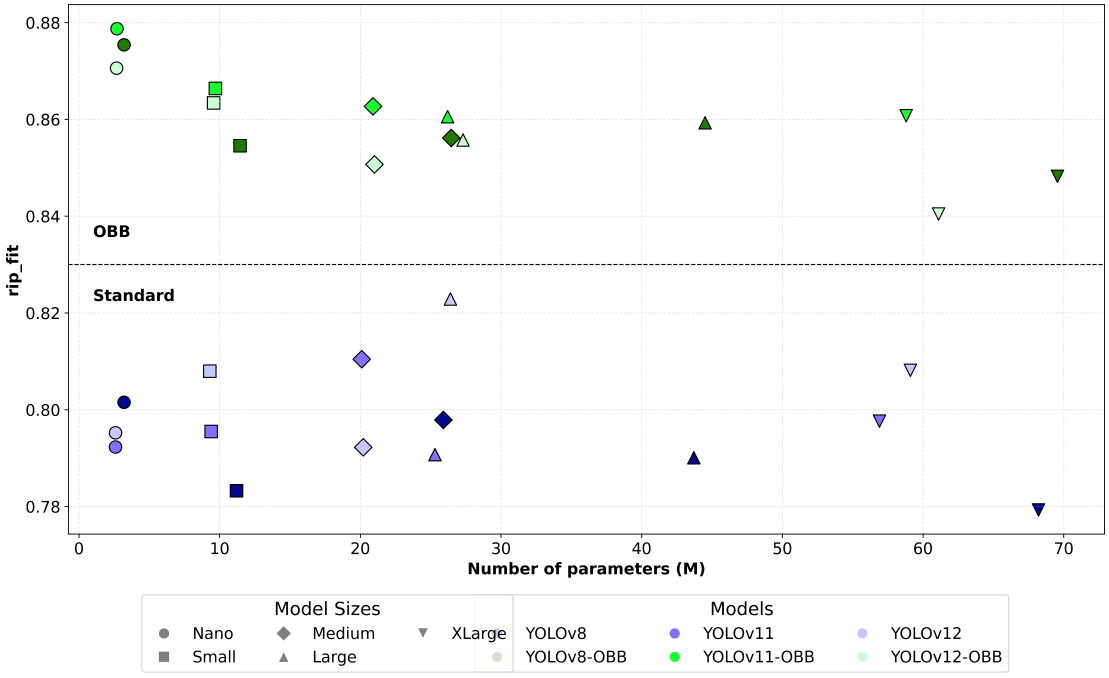

Figure 6. Scatter plot comparison of rip_fit values and number of parameters, for all YOLO model types and sizes. The horizontal line is included to support visual separation of performance metrics based on different annotation setups (*i.e.,* OBBs, standard).

3.3 Class configuration experiment

This experiment assessed the effect of annotation class configuration by comparing conservative (*i.e.,* "doubt" and "sediment" annotations as background) and permissive approaches (*i.e.,* "doubt" and "sediment" categorized as rips). The models used in this experiment were the nano-sized variants of the three YOLO-OBB versions. These variants obtained the highest rip_fit scores (Fig. 6) and were the most parameter-efficient among the OBB models. Figure 7 shows a scatter plot which compares the rip_fit values of these models under conservative versus permissive class configuration. The results obtained using conservative class configurations showed lower and less consistent (more dispersion between models) rip_fit values than their counterparts. The best-performing models for each metric (Table A4), along with their relative changes compared to the annotation type experiment, were: YOLOv11n-OBB for mAP50, $\mathrm{mAP50-95}$, precision and rip_fit (decreases of 11.2%, 25.9%, 12.5% and 10% respectively), and YOLOv8n-OBB for recall (decrease of 5.8%). These results indicated an overall decline in performance across metrics when using a conservative class configuration, compared to the permissive class configuration.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

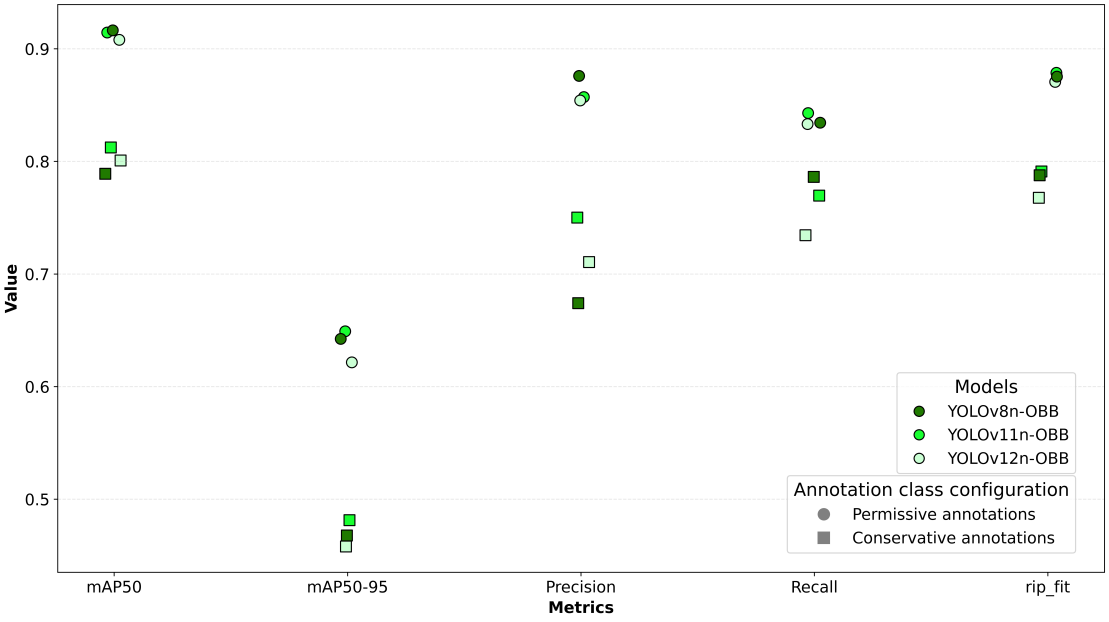

Figure 7. Scatter plot comparing the performance of the models in the third experiment with their performance in the second experiment, across the metrics $\mathrm{mAP50}$, $\mathrm{mAP50-95}$, precision, recall, and $\mathrm{rip\_fit}$.

3.4 Hyperparameter optimization and CrossValidation

Hyperparameter optimization was conducted for the best performance model and annotation type resulting from previous experiments: *i.e.,* YOLOv11n-OBB. Hyperparameter values resulting from the optimization process fell close to default YOLOv11n-OBB settings (Table A5), and yielded negligible performance variability compared to the defaults ($< 1\%$ average deviation). The 10-fold cross-validation exercise demonstrated the robust performance and generalization of the model, resulting in a mean $\mathrm{rip\_fit}$ of $0.873(\pm 0.01)$. The low standard deviation (< 0.02) observed across all metrics (Table 2) confirms the model's stability and reliability, independent of the specific training, validation and test data partitions.

Table 2. Summary of performance metrics for the optimized YOLOv11n-OBB model. Values represent the mean and standard deviation (sd) derived from 10-fold cross-validation. Detailed results for individual folds are provided in Appendix Table A6.

Metric	**Precision**	**Recall**	**mAP50**	**mAP50 – 95**	**rip_fit**
mean±sd	0.873±0.02	0.833±0.01	0.913± 0.01	0.646 ±0.01	0.873 ±0.01

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

3.5 The *socib-rip-currents-detection* module

The optimized YOLO model demonstrated reliable rip current detection performance across diverse acquisition systems, fields of view, and image resolutions (*e.g.,* Fig. 8 b-c, e-g). Quantitatively, the model maintains precision and recall values exceeding 0.7 across a broad range of confidence thresholds (0.1 – 0.7, Fig. 8a), while the Precision-Recall curve confirms sustained accuracy across varying recall levels (Fig. 8d). This performance highlights the model's capability to reliably identify amorphous rip current features despite significant environmental variability. Furthermore, the architecture supports near real-time deployment, achieving inference speeds ranging from 2.4 to 60 FPS depending on the specific hardware configuration (Table 3).

Leveraging this computational efficiency for operational use, we developed the *socib-rip-currents-detection* module, which packages the optimized YOLOv11n model within a containerized Docker environment. Designed as a ready-to-deploy tool, the application is distributed through the iMagine and AI4EOSC marketplaces[3,4]. The complete open-source codebase, including the model weights, is publicly accessible in the corresponding GitHub repository (Oliver-Sansó et al., 2025).

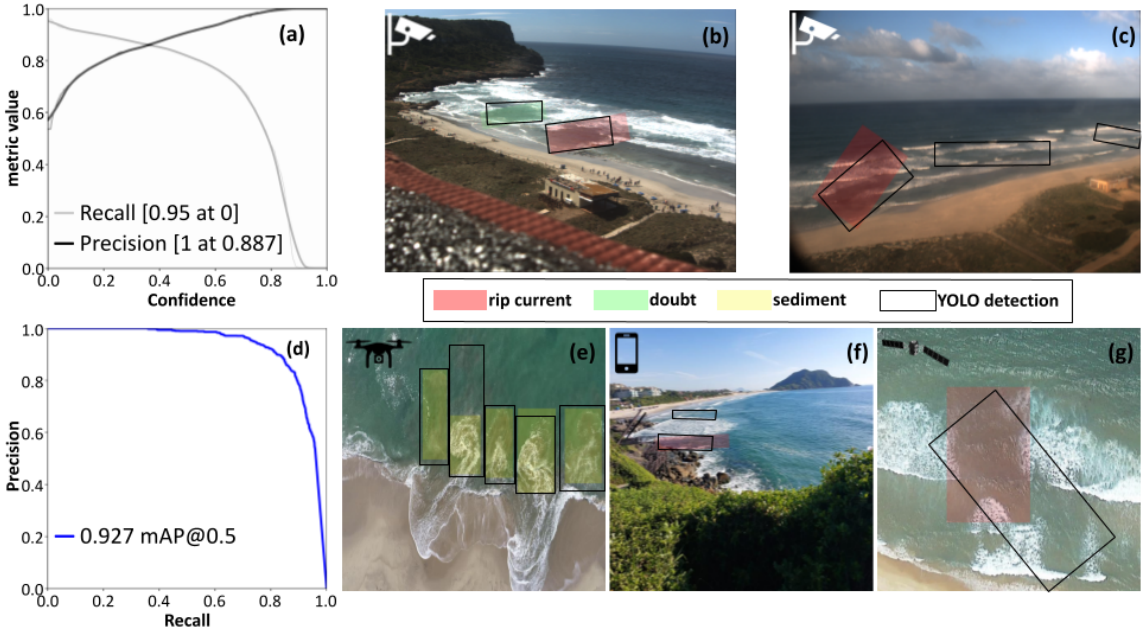

Figure 8. Summarized results for the optimal YOLOv11n-OBB model for rip current detection, evaluated using the test split. (a) Displays recall and precision across confidence thresholds [0-1]. (d) Presents the precision-recall plot computed under IoU of 0.5. Selection of example detections (black rectangles) on various system imagery: (b, c) videomonitoring stations (1280X960 px), (e) Unmanned Aerial Vehicle (640X640 px), (f) smartphone (640X640 px), and (g) satellite (640X640 px). RipAID v2.0.0 annotations are shown with colored rectangles. Note: Ground truth categories are color-coded for reference only; the model was trained treating all label types as a single positive class (permissive strategy).

[3]https://dashboard.cloud.imagine-ai.eu/catalog/modules
[4]https://dashboard.cloud.ai4eosc.eu/catalog/modules

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Table 3. Average inference time (seconds) per image for the optimized YOLOv11n-OBB model, evaluated on the full RipAID v2.0.0 dataset (6,789 images) across six distinct hardware configurations. The time includes pre-processing, prediction, and post-processing, emulating real-world conditions. It is also expressed as Frames Per Second (FPS) for ease of comparison.

System	Computing hardware	Time [s]	speed [FPS]
Type: Raspberry Pi 5; **OS:** Debian GNU/Linux 12 **CPU:** Broadcom BCM2712 (ARM Cortex-A76) 2.4 GHz × 4 **RAM:** 8 GiB; **Storage:** MicroSD	cpu	0.422	2.370
Type: Desktop PC; **OS:** Ubuntu 22.04.5 LTS **CPU:** Intel Xeon(R) CPU E5-2620 v2 2.10 GHz × 12 **RAM:** 16 GiB; **Storage:** SSD	cpu	0.201	4.983
Type: Laptop; **OS:** Ubuntu 24.04.3 LTS **CPU:** 13th Gen Intel Core i7-13700H × 20	cpu	0.224	4.474
RAM: 64 GiB; **Storage:** SSD **Dedicated GPU:** NVIDIA GeForce RTX 3050 Laptop (4 GB)	gpu	0.035	28.902
Type: Desktop Workstation; **OS:** Ubuntu 24.04.3 LTS **CPU:** Intel Xeon w5-3433 × 32	cpu	0.046	21.787
RAM: 64 GiB; **Storage:** SSD **Dedicated GPU:** NVIDIA RTX 5000 Ada Generation (32 GB)	gpu	0.017	60.241

4 Discussion

The primary objective of this study was to advance the automation of rip current detection using beach imaging systems by delivering a robust, platform-agnostic operational tool that integrates novel computer vision approaches. While previous studies have demonstrated the feasibility of deep learning for this task (de Silva et al., 2021; Zhu et al., 2022; Rashid et al., 2023), they have often relied on homogeneous datasets or idealized morphological conditions, resulting in a “generalization gap" when models face the complex realities of diverse coastal environments. By curating RipAID v2.0.0 (Soriano-González et al., 2026), an open-access dataset enriched with multi-platform imagery and oblique viewing angles, this work directly addresses that gap. Beyond varied viewpoints and image geometries, the dataset captures critical rip current characteristics, including images from low-energy coastal environments in the Mediterranean, where rips often exhibit lower intensity and less visible structure, thereby ensuring robust detection across diverse environmental scenarios. The successful performance of the final model on RipAID v2.0.0 demonstrates the efficacy of this approach. As illustrated in Figure 8, the system is capable of processing imagery from diverse sources with varying resolutions and across a range of wave-energy conditions. This includes high-definition video monitoring stations (1280x960 px; Fig. 8b-c) as well as lower-resolution data from UAVs, crowdsourced images collected from smartphones, or satellite imagery (640x640 px; Fig. 8e-g). This capability confirms that training on

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

heterogeneous data allows the model to learn the morphological features of variable rip currents independent of image scale, which is a key prerequisite for operational scalability across image-based coastal monitoring systems.

In the assessment of standard axis-aligned architectures (*i.e.,* use common bounding boxes), FCOS, Faster R-CNN, and RetinaNet demonstrated clear underperformance compared to YOLO and RT-DETR, with the latter two achieving comparable overall scores (Fig. 4). Although these improvements are likely influenced by framework-specific optimizations, such as the default data augmentation pipelines inherent to the Ultralytics ecosystem used for YOLO and RT-DETR, they nonetheless align with findings in the broader literature. Previous studies have similarly identified YOLO-based architectures as optimal for rip current detection (Zhu et al., 2022; Rashid et al., 2023), a trend extending to other research fields dealing with amorphous features (e.g., Geng et al., 2024). However, the most critical factor in achieving robustness proved to be the adoption of Oriented Bounding Boxes (OBBs), representing a fundamental shift away from the axis-aligned paradigms commonly employed in earlier studies (Zhu et al., 2022; de Silva et al., 2021). As amorphous, jet-like hydrodynamic features, rip currents rarely align with the cardinal axes of an image sensor. Consequently, OBBs reduce the inclusion of background noise within the annotated region of a diagonally oriented rip current. This geometric adaptation yielded an average improvement of 12.8% across all metrics compared to axis-aligned counterparts, independent of the YOLO model version or size. Notably, no correlation was observed between increased parameter counts and superior performance (Fig. 6), indicating that for hydro-geomorphological features possessing directional magnitude, the geometric fidelity of the annotation is often more critical than the scale of the model architecture.

Beyond geometric considerations, the definition of the target class itself plays a pivotal role in safety-critical applications. In hazard warning systems, the trade-off between precision and recall is governed by the cost of errors. Given that the envisioned application is a near real-time warning system, minimizing missed detections (false negatives) is more critical than reducing false alarms (false positives). Consequently, the “permissive" annotation strategy, which incorporates ambiguous features labeled as *doubt* or *sediment*, proved superior for operational safety. Notably, YOLO-OBB models demonstrated significant performance gains (ranging from 6% to 26% across metrics) when trained under this permissive configuration compared to conservative baselines (Fig. 7). This approach is validated by the behavior of the model in complex scenarios. Exemplifying this, Figure 8b illustrates a case where the system detects both a clear rip current and a region annotated as *doubt*, effectively handling the label uncertainty. More critically, Figure 8f shows a detection that was initially missed by the human labeler but correctly identified by the model. In an operational context, this tendency to over-detect in uncertain cases serves as a crucial fail-safe mechanism, ensuring that potential ambiguous hazards are flagged for verification. This is supported quantitatively by the optimized YOLOv11n-OBB model, which achieves a recall of 0.889, ensuring that the majority of visible rip currents are captured.

This capacity of the model to occasionally outperform the human annotators highlights a “ground truth paradox” (Plank, 2022). For instance, Figure 8c presents a scenario where the model detected two additional rip currents alongside the labeled instance; post-hoc visual inspection suggests these detections are likely valid. Similarly, Figure 8g depicts a low precision detection where the predicted bounding box diverges from the manual annotation. However, given the fluid and boundary-uncertain nature of rip currents, the model interpretation may be morphologically more accurate than the human label. These

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

instances highlight the subjectivity and complexity of manual annotation, especially in tasks such as rip current identification, which inherently involve an increased level of human label variation. In such tasks, where human agreement can be low, the model may effectively outperform human annotators by learning more stable and consistent pattern than the human labels. This underscores the potential to enhance the ground truth quality, while simultaneously suggesting that the deployed model can serve a dual purpose, not only as a detection system but as a semi-automated dataset refinement or expansion tool. By using the model to pre-annotate new footage or flag inconsistencies in existing datasets, we can accelerate the expansion of future RipAID versions through a human-in-the-loop iterative cycle, capturing complex rip current instances that human annotators may overlook, and expediting the tedious task of annotation without sacrificing human input and expertise.

To integrate the system into coastal monitoring networks, however, accuracy must be matched by computational efficiency. The developed module utilizes the lightweight YOLOv11n-OBB architecture (2.7 M parameters), achieving average inference speeds between 2.4 and 60 FPS (Table 3), depending on hardware configurations (from Raspberry Pi to GPU workstations; Table 3). This efficiency is maintained even in spatially complex scenes. For example, Figure 8e shows the model resolving multiple, closely spaced rip current instances within a single frame, while Figure 8f demonstrates the correct detection of rip currents at varying distances from the camera. These results confirm that the module is sufficiently computationally lightweight to be deployed on edge devices and for incorporation into high-frequency alert systems without creating bottlenecks. Contextually, this performance spectrum covers most of the operational history of coastal imaging: video systems for coastal applications such as shoreline position monitoring or video-derived bathymetry operate at frequencies as low as 2 FPS and up to 7.5 FPS (Holman and Stanley, 2007; Nieto et al., 2010; Simarro et al., 2019), a frequency also applicable to most open surfcam video streams (Andriolo et al., 2019). Applications requiring increased frequency, such as embedded UAV systems, typically use 15-30 FPS (Wang et al., 2025; Hua et al., 2024). Beyond basic system compatibility, this high throughput offers additional strategic advantages: since the model is computationally inexpensive, it can support combination with other techniques, such as multiple-detection voting methods (*e.g.,* detection under varying inference arguments) or temporal consistency based confirmations (*e.g.,* detection persistence along consecutive frames), to further enhance the system's accuracy under complex scenes.

While direct comparison between the results of the proposed model and existing state-of-the-art rip current detection models is challenging due to differences in dataset and evaluation metric definitions, this work represents a distinct advance over existing approaches. Our model successfully generalizes to the oblique angles typical of shore-based cameras, contributing to address previous concerns in the literature (de Silva et al., 2021; Zhu et al., 2022; Rampal et al., 2022; Rashid et al., 2023), and aligning with most recent developments in the field (Khan et al., 2025a). Furthermore, by opting for OBBs, the system captures orientation effectively without the cost of pixel-level masking as in instance segmentation approaches (Dumitriu et al., 2025), easing scalability for future dataset generation.

However, several limitations remain. The system still relies on optical signatures, meaning performance may degrade under poor visibility, unseen environmental conditions, or when rip currents lack surface manifestations (*i.e.,* are not identifiable in images). While the latter is an inherent limitation of optical observation systems, the former presents limitations that can be assessed and mitigated. By recognizing that the coastal environment is not stationary and that concepts learned during

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

training might evolve over time and across imaging systems, tools to evaluate data drift could be employed in order to flag specific images that represent new or challenging data patterns (e.g., Céspedes Sisniega and Álvaro López García, 2024; Azmi et al., 2025a). This enables early identification of performance degradation and supports targeted dataset updates or model retraining, thereby enhancing long-term operational reliability. Additionally, users of the module must consider that while multi-platform, the training dataset does not yet fully cover extreme low-angle geometries, such as those from ground-level smartphone imagery, which could be particularly useful for citizen-oriented initiatives (Khan et al., 2025a). Future work will focus on leveraging the self-refinement capabilities of the model to enhance and expand the training data, with a particular focus on prioritizing the inclusion of underrepresented and ambiguous cases.

Overall, the model achieves robust, autonomous detection capabilities, establishing a new baseline for operational readiness in rip current monitoring. This transition from research to deployment is facilitated by the release of the *socib-rip-currents-detection module* (Oliver-Sansó et al., 2025), an open-source, containerized inference engine designed for immediate integration into multi-platform observation networks. By making the complete codebase and pre-trained weights publicly accessible (Section 3.5), we aim to democratize access to advanced safety tools, enabling the community to re-train, benchmark, and adapt the system to local coastal morphologies without prohibitive development costs. This open ecosystem fosters reproducibility and aligns with broader open science efforts, building upon previous studies and initiatives that have been crucial in the field (de Silva et al., 2021; Zhu et al., 2022; Dumitriu et al., 2025; Khan et al., 2025b, a). Ultimately, its greatest potential lies in the adoption of the module within coastal safety frameworks as a support tool for lifeguards, emergency services, or applications that bridge beach monitoring cameras with end-users. In this role, the system could serve to reduce rip current risk by drawing attention to potential hazards, maximizing coverage while retaining human verification to ensure reliability.

5 Conclusions

This study presents a comprehensive framework for the automated detection of rip currents, moving from dataset curation and model benchmarking to the deployment of an operational solution. A key contribution of this work is the development of RipAID v2.0.0, a novel annotated dataset that significantly expands the diversity of available training data by incorporating multi-platform imagery and a broader range of rip current morphodynamics. This resource addresses the generalization limitations of previous studies, ensuring the model is exposed to the complex realities of diverse coastal environments.

Through systematic benchmarking, the results highlighted the critical role of annotation geometry in detecting amorphous hydrodynamic features. Oriented Bounding Boxes (OBBs) were found to significantly outperform traditional axis-aligned methods, providing a more accurate representation of rip currents within complex nearshore backgrounds. Among the evaluated architectures, the YOLOv11-OBB model emerged as the most effective solution. The nano-sized variant of this architecture (2.7 M parameters) demonstrated a high degree of efficiency, achieving inference speeds capable of supporting real-time applications (2.4–60 FPS). This performance allows the module to be integrated into a wide variety of coastal camera systems, ranging from edge-computing devices to high-frequency video processors, without imposing strict or unaffordable constraints on hardware or image resolution.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

To facilitate the transition from research to practice, the detection framework has been encapsulated into a ready-to-deploy, Dockerized inference module. This provides a tangible operational tool that can be immediately adopted by coastal managers and researchers. Grounded in the principles of open science, the complete codebase, model weights, and the RipAID dataset have been made publicly available. By lowering the technical barriers to entry and fostering reproducibility, this work aims to provide the community with an adaptable foundation for future advancements in AI-driven coastal safety.

Looking forward, future work will focus on addressing the remaining limitations of optical detection. Key priorities include implementing data drift evaluation tools to maintain reliability across diverse imaging systems and expanding the dataset by introducing new scenarios and leveraging the self-refinement capabilities of the model to further capture ambiguous or rare cases. Establishing this continuous learning loop is essential for deploying fully autonomous systems capable of adapting to the evolving dynamics of the coastal environment.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Appendix A

Table A1. Parameters range (minimum and maximum values) and step (interval between possible values) for hyperparameter selection in each Optuna trial, along with their name, default value, and description as defined in the Ultralytics YOLO framework (Jocher et al., 2023b).

Parameter	Default Value	Range	Step	Description
Training hyperparameters				
weight_decay	0.0005	0.0005–0.001	0.0001	L2 regularization term, penalizing large weights to prevent overfitting.
box	7.5	4–7.5	0.5	Weight of the box loss component in the loss function, influencing the emphasis placed on accurate bounding box prediction.
dfl	1.5	1–2	0.25	Weight of the distribution focal loss, used for fine-grained localization in certain YOLO versions.
Data augmentation hyperparameters				
hsv_h	0.015	0.015–0.03	0.05	Adjusts the image hue by a fraction of the color wheel, introducing color variability to improve generalization under varying lighting conditions.
hsv_s	0.7	0.7–0.8	0.05	Modifies image saturation, affecting color intensity and simulating different environmental conditions.
hsv_v	0.4	0.4–0.6	0.05	Adjusts image brightness, enabling robustness across diverse illumination scenarios.
degrees	0.0	−10.0–10.0	2	Randomly rotates images within the specified range to improve robustness to object orientation.
translate	0.1	0.1–0.3	0.1	Translates images horizontally and vertically, aiding detection of partially visible objects.
scale	0.5	0.3–0.6	0.1	Scales images to simulate objects at different distances from the camera.
shear	0.0	0.0–5.0	1	Applies shearing transformations to simulate changes in viewing angle.
flipud	0.0	0.0–0.3	0.1	Flips images vertically with a given probability, increasing data variability.
fliplr	0.5	0.4–0.7	0.1	Flips images horizontally with a given probability, useful for learning symmetric patterns.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Table A2. Metrics for each model in the first experiment. The three best-performing models for each metric are highlighted in bold. In case of a tie, all tied models are highlighted accordingly.

Model Name	mAP50	mAP50–95	Precision	Recall	rip_fit
YOLOv8n	**0.844**	0.467	0.846	0.759	0.802
YOLOv8s	0.826	0.466	0.825	0.741	0.784
YOLOv8m	0.837	0.469	0.843	0.759	0.798
YOLOv8l	0.822	0.460	0.798	0.758	0.790
YOLOv8x	0.831	0.467	**0.864**	0.728	0.780
YOLOv11n	0.839	0.474	0.838	0.746	0.793
YOLOv11s	0.841	0.474	0.844	0.750	0.796
YOLOv11m	0.842	0.472	0.833	**0.779**	**0.811**
YOLOv11l	0.835	**0.475**	0.845	0.746	0.791
YOLOv11x	0.841	**0.477**	0.835	0.754	0.798
YOLOv12n	0.840	0.469	0.832	0.751	0.796
YOLOv12s	0.842	0.471	0.814	0.774	**0.808**
YOLOv12m	**0.844**	**0.475**	**0.862**	0.741	0.793
YOLOv12l	**0.860**	**0.487**	**0.848**	**0.786**	**0.823**
YOLOv12x	**0.847**	**0.475**	0.815	0.769	**0.808**
rtdetr-l	0.816	0.426	0.810	0.757	0.787
rtdetr-x	0.825	0.430	0.846	0.754	0.790
Retinanet	0.683	0.303	0.565	0.767	0.725
FasterRCNN	0.769	0.374	0.545	**0.839**	0.804
FCOS	0.706	0.351	0.697	0.768	0.737

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Table A3. Metrics for each model in the second experiment. The three best-performing models for each metric are highlighted in bold. In case of a tie, all tied models are highlighted accordingly.

Model Name	mAP50	mAP50–95	Precision	Recall	rip_fit
YOLOv8n-OBB	**0.916**	0.642	0.876	**0.834**	**0.875**
YOLOv8s-OBB	0.908	0.639	**0.904**	0.801	0.855
YOLOv8m-OBB	0.896	**0.646**	**0.885**	0.817	0.857
YOLOv8l-OBB	0.896	**0.646**	0.867	0.822	0.859
YOLOv8x-OBB	0.888	0.639	0.881	0.809	0.849
YOLOv11n-OBB	**0.915**	**0.649**	0.857	**0.843**	**0.879**
YOLOv11s-OBB	**0.911**	0.639	0.866	0.821	0.866
YOLOv11m-OBB	0.900	0.644	0.877	0.825	0.863
YOLOv11l-OBB	0.903	**0.652**	0.884	0.818	0.861
YOLOv11x-OBB	0.905	0.643	0.884	0.817	0.861
YOLOv12n-OBB	0.908	0.622	0.854	**0.833**	**0.871**
YOLOv12s-OBB	0.910	0.636	0.865	0.817	0.864
YOLOv12m-OBB	0.910	0.636	0.883	0.792	0.851
YOLOv12l-OBB	0.904	0.636	0.877	0.808	0.856
YOLOv12x-OBB	0.903	0.638	**0.901**	0.778	0.841

Table A4. Metrics for each model in the third experiment. The best-performing model for each metric is highlighted in bold.

Model Name	mAP50	mAP50–95	Precision	Recall	rip_fit
YOLOv8n-OBB	0.789	0.468	0.674	**0.786**	0.788
YOLOv11n-OBB	**0.812**	**0.481**	**0.750**	0.770	**0.791**
YOLOv12n-OBB	0.801	0.458	0.711	0.734	0.768

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Table A5. Hyperparameter optimization results. Default values are included for ease of comparison. The three most important parameters are highlighted in bold.

Parameter	Default Value	Optimized Value	Param Importance
Training hyperparameters			
weight_decay	0.0005	0.0009	0.024
box	7.5	6.5	0.055
dfl	1.5	1.0	**0.458**
Data augmentation hyperparameters			
hsv_h	0.015	0.03	0.028
hsv_s	0.7	0.7	**0.065**
hsv_v	0.4	0.45	0.018
degrees	0.0	-2	0.047
translate	0.1	0.2	0.013
scale	0.5	0.3	**0.147**
shear	0.0	0	0.055
flipud	0.0	0	0.029
fliplr	0.5	0.4	0.061

Table A6. Detailed results of performance metrics for the optimized YOLOv11n-OBB model for individual folds of the cross-validation

Fold	mAP50	mAP50–95	Precision	Recall	rip_fit
Fold 0	0.920	0.651	0.892	0.825	0.872
Fold 1	0.907	0.646	0.851	0.841	0.874
Fold 2	0.910	0.660	0.867	0.826	0.868
Fold 3	0.905	0.650	0.897	0.816	0.860
Fold 4	0.911	0.628	0.879	0.823	0.867
Fold 5	0.917	0.633	0.879	0.846	0.882
Fold 6	0.927	0.677	0.888	0.845	0.886
Fold 7	0.906	0.639	0.842	0.837	0.872
Fold 8	0.909	0.634	0.858	0.839	0.874
Fold 9	0.920	0.639	0.873	0.832	0.876

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Code and data availability. The *socib-rip-currents-detection* module is distributed through the iMagine[5] and AI4EOSC[6] marketplaces. The codebase and pre-trained weights are hosted in the AI4OS-hub repository[7]. The RipAID v2.0.0 dataset used for model training, along with an associated README file detailing dataset generation and metadata, is available in the Zenodo open access repository (Soriano-González et al., 2026).

Author contributions. Conceptualization, Data curation, Investigation, Methodology and Writing (original draft preparation): A.C-G and J.S-G; Funding acquisition: A.F-M and F.C-S; Resources: V.K and K.A; Software: A.C-G, J.S-G, J.O-S and K.A; Supervision: J.S-G, A.F-M and J.L; Visualization: A.C-G, J.S-G and E.S-G. All co-authors participated in Writing (review and editing)

Competing interests. The authors declare no competing interests.

Acknowledgements. This work was supported by the EU project 'iMagine: Imaging data and services for aquatic science' with funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101058625. A. Català-Gonell was supported by a SOCIB grant for an extracurricular MSc internship. Authors acknowledge the support of the Spanish Ministry of Science, Innovation and Universities, the Government of the Balearic Islands, and the Spanish National Research Council (CSIC), as well as the support of the General Directorate of Emergencies and Home Affairs of the Balearic Islands (Direcció General d'Emergències i Interior) and the lifeguard teams of the Balearic Islands for their collaboration in data collection. Finally, we also extend our gratitude to the authors of the public datasets and open-source initiatives cited in this manuscript, whose contributions were fundamental to the development of this work.

During the preparation of this manuscript, the authors utilized Large Language Model tools to assist with language editing and to improve readability. These tools were not used for data generation, analysis, or the scientific interpretation of the results, and the authors take full responsibility for the final content of this publication.

[5]https://dashboard.cloud.imagine-ai.eu/catalog/modules/socib-rip-currents-detection
[6]https://dashboard.cloud.ai4eosc.eu/catalog/modules/socib-rip-currents-detection
[7]https://github.com/ai4os-hub/socib-rip-currents-detection

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

References

CnE UFSC : Rip Current - coastsnap - Moçambique + Santinho Dataset, https://universe.roboflow.com/cne-ufsc/rip-current-coastsnap-mocambique-santinho, [Data set], 2023.

Adade, R., Aibinu, A. M., Ekumah, B., and Asaana, J.: Unmanned Aerial Vehicle (UAV) applications in coastal zone management—a review, Environmental Monitoring and Assessment, pp. 154–193, https://doi.org/https://doi.org/10.1007/s10661-021-08949-8, 2019.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, p. 2623–2631, Association for Computing Machinery, New York, NY, USA, ISBN 9781450362016, https://doi.org/10.1145/3292500.3330701, 2019.

Andriolo, U., Sánchez-García, E., and Taborda, R.: Operational Use of Surfcam Online Streaming Images for Coastal Morphodynamic Studies, Remote Sensing, 11, https://doi.org/10.3390/rs11010078, 2019.

Azmi, E., Alibabaei, K., Kozlov, V., Krijger, T., Accarino, G., Ayata, S.-D., Calatrava, A., De Carlo, M. M., Decrop, W., Elia, D., Fiore, S. L., Francescangeli, M., Irisson, J.-O., Lagaisse, R., Laviale, M., Lebeaud, A., Leluschko, C., Martínez, E., Moltó, G., Ruiz Atake, I., Sepp Neves, A. A., Smyth, D., Soriano-González, J., Tayyab, M. A., Tosello, V., Álvaro López García, Schaap, D., and Sipos, G.: Best practices for AI-based image analysis applications in aquatic sciences: The iMagine case study, Ecological Informatics, 91, 103 306, https://doi.org/10.1016/j.ecoinf.2025.103306, 2025a.

Azmi, E., Alibabaei, K., Kozlov, V., López García, A., Schaap, D., and Sipos, G.: iMagine: AI-Powered Image Data Analysis in Aquatic Science, in: Proceedings of the Platform for Advanced Scientific Computing Conference, PASC '25, p. 1–11, Association for Computing Machinery, New York, NY, USA, ISBN 9798400718861, https://doi.org/10.1145/3732775.3733584, 2025b.

Bowman, D., Arad, D., Rosen, D., Kit, E., Goldbery, R., and Slavicz, A.: Flow characteristics along the rip current system under low-energy conditions, Marine Geology, 82, 149–167, https://doi.org/10.1016/0025-3227(88)90138-7, 1988.

Brander, R. and Scott, T.: Science of the rip current hazard, in: The science of beach lifeguarding, pp. 67–85, CRC Press, https://doi.org/10.4324/9781315371641-5, 2018.

Caldwell, N., Houser, C., and Meyer-Arendt, K.: Ability of beach users to identify rip currents at Pensacola Beach, Florida, Natural Hazards, 68, 1041–1056, https://doi.org/10.1007/s11069-013-0673-3, 2013.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S.: End-to-End Object Detection with Transformers, in: Computer Vision – ECCV 2020, edited by Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., pp. 213–229, Springer International Publishing, Cham, ISBN 978-3-030-58452-8, https://doi.org/10.1007/978-3-030-58452-8_13, 2020.

Castelle, B., Scott, T., Brander, R., and McCarroll, R.: Rip current types, circulation and hazard, Earth-Science Reviews, 163, 1–21, https://doi.org/10.1016/j.earscirev.2016.09.008, 2016.

CVAT.ai Corporation: Computer Vision Annotation Tool (CVAT), https://doi.org/10.5281/zenodo.7863887, 2023.

Céspedes Sisniega, J. and Álvaro López García: Frouros: An open-source Python library for drift detection in machine learning systems, SoftwareX, 26, 101 733, https://doi.org/https://doi.org/10.1016/j.softx.2024.101733, 2024.

de Silva, A., Mori, I., Dusek, G., Davis, J., and Pang, A.: Automated rip current detection with region based convolutional neural networks, Coastal Engineering, 166, 103 859, https://doi.org/10.1016/j.coastaleng.2021.103859, 2021.

Dumitriu, A., Tatui, F., Miron, F., Ionescu, R. T., and Timofte, R.: Rip Current Segmentation: A Novel Benchmark and YOLOv8 Baseline Results, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1261–1271, https://doi.org/10.1109/CVPRW59228.2023.00133, 2023.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Dumitriu, A., Tatui, F., Miron, F., Ralhan, A., Ionescu, R. T., and Timofte, R.: RipVIS: Rip Currents Video Instance Segmentation Benchmark for Beach Monitoring and Safety, in: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3427–3437, https://doi.org/10.1109/CVPR52734.2025.00325, 2025.

Dérian, P. and Almar, R.: Wavelet-Based Optical Flow Estimation of Instant Surface Currents From Shore-Based and UAV Videos, IEEE Transactions on Geoscience and Remote Sensing, 55, 5790–5797, https://doi.org/10.1109/TGRS.2017.2714202, 2017.

Geng, X., Su, Y., Cao, X., Li, H., and Liu, L.: Scientific Reports, Scientific Reports, 14, https://doi.org/10.1038/s41598-024-55232-0, 2024.

Heredia, I., Álvaro López García, Moltó, G., Calatrava, A., Kozlov, V., Costantini, A., Tran, V., David, M., Martín, D. S., Płóciennik, M., Ruiz, M. O., Fernandez, S., Díaz, J. S.-P., Caballer, M., Marín, C. A., Dlugolinsky, S., Šeleng, M., Berberi, L., Alibabaei, K., Sanchis, B. E., Castro, P., Donvito, G., Aguirre, D., Langarita, S., Rodriguez, V., Duda, L., Canales, A. H., Ruiz, S. R., Machado, J., Nguyen, G., Gómez, F. A., and Díez, J.: AI4EOSC: a Federated Cloud Platform for Artificial Intelligence in Scientific Research, https://arxiv.org/abs/2512.16455, 2025.

Holman, R. and Stanley, J.: The history and technical capabilities of Argus, Coastal Engineering, 54, 477–491, https://doi.org/10.1016/j.coastaleng.2007.01.003, the CoastView Project: Developing coastal video monitoring systems in support of coastal zone management, 2007.

Holman, R. A., Symonds, G., Thornton, E. B., and Ranasinghe, R.: Rip spacing and persistence on an embayed beach, Journal of Geophysical Research: Oceans, 111, https://doi.org/10.1029/2005JC002965, 2006.

Hua, W., Chen, Q., and Chen, W.: A new lightweight network for efficient UAV object detection, Scientific Reports, 14, https://doi.org/10.1038/s41598-024-64232-z, 2024.

Huntley, D. A., Hendry, M. D., Haines, J., and Greenidge, B.: Waves and Rip Currents on a Caribbean Pocket Beach, Jamaica, Journal of Coastal Research, 4, 69–79, http://www.jstor.org/stable/4297373, 1988.

Jayagopal, P., Purushothaman Janaki, K., Mohan, P., Kondapaneni, U. B., Periyasamy, J., Mathivanan, S. K., and Dalu, G. T.: A modified generative adversarial networks with Yolov5 for automated forest health diagnosis from aerial imagery and Tabu search algorithm, Scientific Reports, 14, https://doi.org/10.1038/s41598-024-54399-w, 2024.

Jocher, G. and Qiu, J.: Ultralytics YOLO11, https://github.com/ultralytics/ultralytics, 2024.

Jocher, G., Chaurasia, A., and Qiu, J.: Ultralytics YOLOv8, https://github.com/ultralytics/ultralytics, 2023a.

Jocher, G., Qiu, J., and Chaurasia, A.: Ultralytics YOLO, https://github.com/ultralytics/ultralytics, 2023b.

Khan, F., de Silva, A., Palinkas, A., Dusek, G., Davis, J., and Pang, A.: RipFinder: real-time rip current detection on mobile devices, Frontiers in Marine Science, Volume 12 - 2025, https://doi.org/10.3389/fmars.2025.1549513, 2025a.

Khan, F. H., Stewart, D., de Silva, A., Palinkas, A., Dusek, G., Davis, J., and Pang, A.: RipScout: Realtime ML-Assisted Rip Current Detection and Automated Data Collection Using UAVs, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 7742–7755, https://doi.org/10.1109/JSTARS.2025.3543695, 2025b.

Khan, Z. F., Ramzan, M., Raza, M., Khan, M. A., Alasiry, A., Marzougui, M., and Shin, J.: Real-Time Polyp Detection From Endoscopic Images Using YOLOv8 With YOLO-Score Metrics for Enhanced Suitability Assessment, IEEE Access, 12, 176 346–176 362, https://doi.org/10.1109/ACCESS.2024.3505619, 2024.

Li, Y., Xie, S., Chen, X., Dollar, P., He, K., and Girshick, R.: Benchmarking Detection Transfer Learning with Vision Transformers, https://arxiv.org/abs/2111.11429, 2021.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal Loss for Dense Object Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42, 318–327, https://doi.org/10.1109/TPAMI.2018.2858826, 2020.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Lippmann, T. C. and Holman, R. A.: Quantification of sand bar morphology: A video technique based on wave dissipation, Journal of Geophysical Research: Oceans, 94, 995–1011, https://doi.org/10.1029/JC094iC01p00995, 1989.

Liu, Y. and Wu, C. H.: Lifeguarding Operational Camera Kiosk System (LOCKS) for flash rip warning: Development and application, Coastal Engineering, 152, 103 537, https://doi.org/10.1016/j.coastaleng.2019.103537, 2019.

Lv, Z., Wang, R., Wang, Y., Zhou, F., and Guo, N.: Road Scene Multi-Object Detection Algorithm Based on CMS-YOLO, IEEE Access, 11, 121 190–121 201, https://doi.org/10.1109/ACCESS.2023.3327735, 2023.

MacMahan, J. H., Thornton, E. B., Stanton, T. P., and Reniers, A. J.: RIPEX: Observations of a rip current system, Marine Geology, 218, 113–134, https://doi.org/10.1016/j.margeo.2005.03.019, 2005.

MacMahan, J. H., Thornton, E. B., and Reniers, A. J.: Rip current review, Coastal Engineering, 53, 191–208, https://doi.org/10.1016/j.coastaleng.2005.10.009, 2006.

Mori, I., de Silva, A., Dusek, G., Davis, J., and Pang, A.: Flow-Based Rip Current Detection and Visualization, IEEE Access, 10, 6483–6495, https://doi.org/10.1109/ACCESS.2022.3140340, 2022.

Nieto, M. A., Garau, B., Balle, S., Simarro, G., Zarruk, G. A., Ortiz, A., Tintoré, J., Álvarez Ellacuría, A., Gómez-Pujol, L., and Orfila, A.: An open source, low cost video-based coastal monitoring system, Earth Surface Processes and Landforms, 35, 1712–1719, https://doi.org/10.1002/esp.2025, 2010.

Oliver-Sansó, J., Soriano-González, J., and Català-Gonell, A.: ai4os-hub/socib-rip-currents-detection: v1.0.0, https://doi.org/10.5281/zenodo.18254942, 2025.

Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., and da Silva, E. A. B.: A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit, Electronics, 10, https://doi.org/10.3390/electronics10030279, 2021.

Pitman, S., Gallop, S. L., Haigh, I. D., Mahmoodi, S., Masselink, G., and Ranasinghe, R.: Synthetic Imagery for the Automated Detection of Rip Currents, Journal of Coastal Research, 75, 912 – 916, https://doi.org/10.2112/SI75-183.1, 2016.

Plank, B.: The 'Problem' of Human Label Variation: On Ground Truth in Data, Modeling and Evaluation, https://arxiv.org/abs/2211.02570, 2022.

Rampal, N., Shand, T., Wooler, A., and Rautenbach, C.: Interpretable Deep Learning Applied to Rip Current Detection and Localization, Remote Sensing, 14, https://doi.org/10.3390/rs14236048, 2022.

Rashid, A. H., Razzak, I., Tanveer, M., and Hobbs, M.: Reducing rip current drowning: An improved residual based lightweight deep architecture for rip detection, ISA Transactions, 132, 199–207, https://doi.org/10.1016/j.isatra.2022.05.015, 2023.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, https://doi.org/10.1109/CVPR.2016.91, 2016.

Ren, S., He, K., Girshick, R., and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031, 2017.

Short, A.: Rip-current type, spacing and persistence, Narrabeen Beach, Australia, Marine Geology, 65, 47–71, https://doi.org/10.1016/0025-3227(85)90046-5, 1985.

Short, A. D.: Australian Rip Systems – Friend or Foe?, Journal of Coastal Research, 50, 7, https://doi.org/10.2112/JCR-SI50-002.1, 2007.

Simarro, G., Calvete, D., Luque, P., Orfila, A., and Ribas, F.: UBathy: A New Approach for Bathymetric Inversion from Video Imagery, Remote Sensing, 11, https://doi.org/10.3390/rs11232722, 2019.

Song, D., Kim, G., and Lee, J.: Quantitative Behaviour Characteristics Analysis of GPS Buoy in Rip Current Generation, Journal of Coastal Research, 75, 1242 – 1246, https://doi.org/10.2112/SI75-249.1, 2016.

https://doi.org/10.5194/egusphere-2026-1138
Preprint. Discussion started: 20 March 2026

Soriano-González, J., Català-Gonell, A., González-Pérez, L., Criado-Sudau, F., Sánchez-García, E., Fernández-Mora, Á., and SOCIB: RipAID: Rip current Annotated Image Dataset (1.0.0), https://doi.org/10.5281/zenodo.15082427, [Data set], 2025.

Soriano-González, J., Català-Gonell, A., González-Pérez, L., Criado-Sudau, F., Sánchez-García, E., Fernández-Mora, Á., and SOCIB: RipAID: Rip Current Annotated Image Dataset (2.0.0), https://doi.org/10.5281/zenodo.18196300, [Dataset], 2026.

Soriano-González, J., Sánchez-García, E., and González-Villanueva, R.: From a citizen science programme to a coastline monitoring system: Achievements and lessons learnt from the Spanish CoastSnap network, Ocean & Coastal Management, 256, 107 280, https://doi.org/10.1016/j.ocecoaman.2024.107280, 2024.

Surf Life Saving Australia: National Coastal Safety Report, https://www.surflifesaving.com.au/wp-content/uploads/sites/2/2021/07/SLSA-National-Coastal-Safety-Report-2019-2.pdf, 2019.

Tian, Y., Ye, Q., and Doermann, D.: YOLOv12: Attention-Centric Real-Time Object Detectors, https://arxiv.org/abs/2502.12524, 2025.

Tian, Z., Shen, C., Chen, H., and He, T.: FCOS: Fully Convolutional One-Stage Object Detection, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9626–9635, https://doi.org/10.1109/ICCV.2019.00972, 2019.

Wang, H.-M., Doong, D.-J., and Lai, J.-W.: Rip Current Identification in Optical Images Using Wavelet Transform, Journal of Marine Science and Engineering, 13, https://doi.org/10.3390/jmse13040707, 2025.

Zaharia, M. A., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Xie, F., and Zumar, C.: Accelerating the Machine Learning Lifecycle with MLflow, IEEE Data Engineering Bulletin, 41, 39–45, https://api.semanticscholar.org/CorpusID:83459546, 2018.

Zhang, S., Chi, C., Yao, Y., Lei, Z., and Li, S. Z.: Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9756–9765, https://doi.org/10.1109/CVPR42600.2020.00978, 2020.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., and Chen, J.: DETRs Beat YOLOs on Real-time Object Detection, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16 965–16 974, https://doi.org/10.1109/CVPR52733.2024.01605, 2024.

Zhu, D., Qi, R., Hu, P., Su, Q., Qin, X., and Li, Z.: YOLO-Rip: A modified lightweight network for Rip currents detection, Frontiers in Marine Science, 9, https://doi.org/10.3389/fmars.2022.930478, 2022.

Zhu, H., Wei, H., Li, B., Yuan, X., and Kehtarnavaz, N.: A Review of Video Object Detection: Datasets, Metrics and Methods, Applied Sciences, 10, https://doi.org/10.3390/app10217834, 2020.