Multi-Source Remote Sensing for Maritime Security: A Performance Evaluation of SAR and RGB Imagery for Small-Scale Fishing Vessel Detection

Stefano Chiappini[1], Alessandro Galdelli[2,4], Andrea Fiorani[2] Marsia Sanità[1], Carmen Ferrà[3,4], Anna Nora Tassetti[3,4], Roberto Pierdicca[1]

[1] Department of Civil, Building Engineering and Architecture (DICEA), Università Politecnica delle Marche, 60131 Ancona, Italy - s.chiappini@staff.univpm.it, m.sanita@pm.univpm.it, r.pierdicca@staff.univpm.it
[2] Department of Information Engineering (D3A), Università Politecnica delle Marche, 60131 Ancona, Italy - a.galdelli@staff.univpm.it, a.fiorani@studenti.univpm.it
[3] CNR-IRBIM, Institute for Marine Biological Resources and Biotechnology, National Research Council, 60125 Ancona, Italy - annanora.tassetti@cnr.it, carmen.ferravega@cnr.it
[4] NBFC, National Biodiversity Future Center, 90133, Palermo, Italy

Keywords: Maritime Surveillance, Deep Learning, Remote Sensing, Small-Scale Fisheries, Vessel Detection.

Abstract

Effective maritime surveillance and small-scale fisheries management remain challenging in coastal waters, where small vessels are not systematically tracked and are often poorly represented in medium-resolution satellite imagery. Within the AI4COPSEC Horizon Europe framework, this study investigates an object-detection workflow for monitoring small vessels along the Adriatic coasts of Marche and Puglia, Italy. Sentinel-2 and high-resolution PlanetScope RGB imagery were manually annotated to build a task-specific optical dataset and to fine-tune models previously pretrained on a larger SAR-optical vessel dataset. This two-stage strategy was designed to exploit heterogeneous vessel representations during pretraining and then adapt the detector to the target coastal optical domain. The resulting dataset comprised 4,202 image tiles for pretraining and 706 tiles for fine-tuning, with 16,096 and 1,716 vessel annotations, respectively, all belonging to a single target class. Detection experiments were conducted using several YOLOv26 configurations trained under a consistent protocol to assess the trade-off between model complexity, accuracy, and computational efficiency. Among the standard variants, YOLOv26-M achieved the most balanced performance, with Precision of 0.813, Recall of 0.846, F1-score of 0.829, Accuracy of 0.719 and mAP50-95 of 0.306. Pruned and lightweight alternatives showed competitive efficiency-oriented behaviour. Results indicate that, in small-target coastal environments, increasing model size does not necessarily yield proportional gains, whereas task-oriented architectural design improves the balance between detection quality and computational cost.

1. Introduction

The monitoring of maritime activities plays an important role in ocean governance, supporting navigation safety, environmental protection, and the sustainable management of marine resources Christodoulou and Echebarria Fernández (2021). Among these activities, the monitoring of fishing operations remains particularly challenging. Illegal, unreported and unregulated (IUU) fishing continues to undermine fisheries management, affecting both marine ecosystems and the economic stability of coastal communities Sumaila et al. (2020); Orofino et al. (2023). Within European waters, fisheries management is regulated under the Common Fisheries Policy (CFP), established by Regulation (EU) No. 1380/2013 (European Parliament and Council of the European Union, 2013), which defines the framework for fisheries conservation and management across EU waters. These mechanisms rely on a combination of reporting systems and technological tools designed to track vessel activity and fishing effort Orofino et al. (2023). However, monitoring capabilities are not uniform across the fishing fleet. European large scale fishing vessels are equipped with tracking systems, most notably Vessel Monitoring Systems (VMS) and Automatic Identification Systems (AIS) that provide regular information on their position enabling detailed reconstructions of fishing footprints and operational patterns Shepperson et al. (2018). In contrast, small-scale fisheries (SSF), defined in the EU as vessels under 12 meters in length that do not use towed gears, remain significantly challenging to monitor Lattanzi et al. (2024). These small vessels operate within 3 nautical miles from the shoreline and are not systematically tracked through cooperative monitoring systems and their activities reported. This monitoring gap has contributed to recent policy developments, including Regulation (EU) 2023/2842 (European Parliament and Council of the European Union, 2023), which extends tracking requirements to small-scale fishing vessels and has motivated increasing interest in non-cooperative monitoring approaches. However, conventional EO approaches face fundamental constraints Galdelli et al. (2020). In maritime surveillance, vessel detection is constrained by a fundamental trade-off between spatial resolution, weather independence and revisit frequency. This limitation is particularly critical for monitoring small vessels in coastal waters, posing a significant challenge for the advancement of Copernicus-based security services. Synthetic Aperture Radar (SAR) sensors, such as Sentinel-1 European Space Agency (2026b), constitute a fundamental pillar for continuous observation, supporting all-weather, day-and-night acquisitions regardless of cloud cover Guan et al. (2025). Complementing this, open-access multispectral sensors such as Sentinel-2 European Space Agency (2026c) offer rich visual and spectral insights. However, their operational utility is strictly limited by atmospheric conditions; furthermore, the medium spatial resolution of freely available Sentinel-2 data is often insufficient to reliably detect small recreational or artisanal fishing vessels. Such targets frequently remain weakly represented or entirely unresolved in medium-resolution imagery La and Pham (2024). Within this framework, high-resolution commercial constellations such as PlanetScope Planet Labs PBC (2024) present a pivotal opportunity to enhance coastal monitoring. By combining high spatial resolution with near-daily revisit capabilities, PlanetScope provides an observational scale better suited for detecting small maritime targets in nearshore environments. Consequently, the integration and comparison of open-access data with higher-resolution commercial imagery represents a promising approach to assess how complementary EO sources can improve vessel detection in complex coastal scenarios Galdelli et al. (2025). Recent advances in Deep Learning, particularly YOLO architectures, have transformed object detection in remote sensing applications. YOLO algorithms achieve compelling speed-accuracy trade-offs that traditional methods cannot match Wang et al. (2019). Researchers have successfully adapted these models to remote sensing, optimizing feature extraction to handle multi-scale variations and the inherent challenge of small target detection Zhang et al. (2021a). Parallel progress in sensor fusion has shown promise because combining SAR, optical imagery and AIS data streams has measurably improved detection robustness for medium to large vessel categories Li et al. (2025). Yet a critical gap remains between promising research results and operational reality. While recent work has begun integrating auxiliary data sources to refine classification performance Galdelli et al. (2021), the effectiveness of fusion approaches remains constrained by open-access sensor resolution. Current research has not systematically addressed how to operationally integrate high-resolution commercial imagery with Copernicus open-access products for the detection and classification of SSF vessels. Validated, integrated frameworks capable of bridging this methodological divide remain absent from the literature. This operational gap constitutes the precise focal point of the research presented herein. The present study is consistent with the objectives of the Horizon Europe AI4COPSEC project Anderson et al. (2025), which aims to strengthen the Copernicus Security Service through AI-based solutions for maritime surveillance, anomaly detection and IUU fishing monitoring. In this context, the research addresses the recognition of unreported SSF vessels operating close to the coast without cooperative tracking systems, which constitute a particularly challenging target class for maritime security and fishery management applications. Building on a previous study by the same research group, which integrated Sentinel-1 SAR, Sentinel-2, high-resolution optical imagery, and AIS data to improve vessel detection and identification of potentially suspicious maritime activities Galdelli et al. (2025), the present work extends the analysis by specifically focusing on small fishing vessels. To this end, a dedicated dataset comprising Sentinel-2 and high-resolution PlanetScope optical imagery was constructed and manually annotated, and the experimental analysis was restricted to RGB optical data which provide the spatial and spectral detail required to resolve small targets. A two-stage training protocol was adopted to improve representation learning and domain adaptation. Models were first pretrained on a larger, SAR and optical dataset to capture generalizable visual features, and subsequently fine-tuned on the curated SSF-specific dataset to enhance sensitivity to small-object characteristics and optimize detection performance under domain-specific conditions. These assessments are conducted under the operational constraints of near-shore monitoring within 3 nautical miles of the coast. To meet the stringent latency requirements of near real-time surveillance, the study adopts the high-efficiency YOLOv26 architecture as the core detection model. Overall, the study contributes to the development of a benchmark for small vessels detection and supports the advancement of more effective strategies against IUU fishing activities.

2. Materials and methodology

2.1 Study Area

The study focuses on the coastal waters of the Adriatic Sea, specifically bordering the Marche and Puglia regions in Italy (Figure 1). Case study areas are defined by reference coordinates (13.489° E, 43.528° N and 18.014° E, 40.584° N, respectively) and extend approximately 3 nautical miles from the coastline, representing a critical domain for SSF and nearshore maritime traffic.

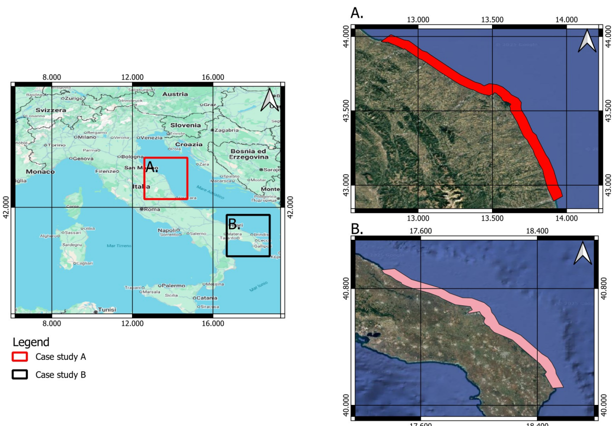


Figure 1. Study areas in the Adriatic Sea. **Left:** Overview of the two study areas along the Italian coastline. **Top right (A)**: Detailed view of the Marche region. **Bottom right (B)**: Detailed view of the Puglia region. The colored polygons outline the coastal belt extending up to 3 nautical miles from the shoreline.

The investigated coastal environment is characterized by a heterogeneous nearshore setting, including proximity to the coastline, port and harbor infrastructure, and seasonally variable maritime traffic. In accordance with Article 13 of EU Council Regulation 1967/2006[1], the coastal belt extending up to 3 nautical miles from the shoreline, or to the 50 m isobath where this occurs closer to shore, represents a zone where small-scale fishing activities are prevalent due to restrictions on towed gears. These characteristics make the selected sectors particularly suitable for assessing small-vessel detection in coastal and non-cooperative maritime scenarios. The regulatory framework also highlights the need for increased monitoring to support effective nearshore management and compliance.

2.2 Dataset Preparation

With the aim to focus on SSF activities off the coasts, Sentinel-2 scenes were downloaded from the Copernicus Data Space Ecosystem[2], while PlanetScope imagery was retrieved from the Planet Explorer platform[3]. For each sensor and each case

[1] Council Regulation (EC) No. 1967/2006 concerning management measures for the sustainable exploitation of fishery resources in the Mediterranean Sea, amending Regulation (EEC) No. 2847/93 and repealing Regulation (EC) No. 1626/94.

[2] `https://browser.dataspace.copernicus.eu/`, accessed on 1 September 2025.

[3] `https://www.planet.com/explorer/`, accessed on 1 September 2025.

study area, eight images acquired between early June and late September 2024 were selected. This time window corresponds to the summer season and overlaps with the seasonal closure of Italian large-scale fisheries in the selected regions. To ensure optimal optical visibility, only acquisitions with 0% cloud cover were considered.

Sentinel-2 imagery was adopted as the medium-resolution optical component. Level-2A Bottom-of-Atmosphere (BOA) products were used, and considered analysis-ready because they are generated from Level-1C data through the Sen2Cor processor European Space Agency (2026c). This provides atmospherically corrected surface reflectance products and associated quality layers, ensuring radiometric consistency and comparability across acquisition dates without the need for additional atmospheric preprocessing. To ensure consistency in subsequent analysis, only the RGB bands were retained, namely B02 (blue, 490 nm), B03 (green, 560 nm) and B04 (red, 665 nm) European Space Agency (2026a),with a ground sampling distance (GSD) ranging from 10 m. PlanetScope imagery was used as the high-resolution optical component of the dataset, specifically PlanetScope Scene products (PSScene) acquired by the PSB.SD SuperDove sensors. According to the Planet product specification, these products consist of individual framed scenes acquired within an image strip and provide imagery with a nominal pixel size of 3 m and a ground sample distance of approximately 3–4 m Planet Labs PBC (2025). Although PSB.SD products provide eight spectral bands, only the visible RGB channels were retained, corresponding to blue (490 nm, FWHM 50 nm), green (565 nm, FWHM 36 nm), and red (665 nm, FWHM 31 nm). These bands are explicitly designed to be interoperable with the corresponding Sentinel-2 visible bands, namely B02, B03, and B04, thus supporting a direct cross-sensor comparison Planet Labs PBC (2024). As illustrated in Figure 2, the preprocessed raster data were imported into QGIS, where a regular, non-overlapping grid was applied to partition the large geospatial imagery into manageable patches. Specifically, a tile size of 640×640 pixels was adopted. This dimension was selected to balance computational resource constraints with the need to adequately capture the small physical size of the targets of interest. To streamline the annotation phase, a targeted extraction strategy was adopted. Tiles containing candidate vessel instances were isolated and exported in `.jpg` format. The resulting collection of selected tiles was then uploaded to the Label Studio platform Tkachenko et al. (2022), where manual annotation was performed by drawing rectangular bounding boxes around each visible vessel. Since the detection task was formulated as a single-class problem, all annotated instances were assigned to the label ”vessel”. The dataset and corresponding annotations were then exported in standard YOLO format, with bounding box coordinates normalized to the image dimensions to ensure compatibility with the subsequent training pipeline. Table 1 summarizes the total number of manually annotated objects.

Table 1. Total number of vessels annotated by sensor, month, and case study area. Case study A refers to the Marche Region, whereas case study B refers to the Puglia Region.

Sensor	June		July		August		September	
	A	B	A	B	A	B	A	B
Sentinel-2	45	12	137	223	50	26	19	1
PlanetScope	242	53	172	42	361	120	41	172

Specifically, three publicly available datasets were leveraged for initial pretraining: (i) the SAR Ship Detection Dataset (SSDD) Zhang et al. (2021b) containing 1,160 images (812 trainig, 174 validation, and 174 testing); and (ii) two Sentinel-2 datasets, namely the Sentinel-2 Ship Detection Dataset (S2_DE) Roboflow Universe (2023) containing 1,747 images (1,225 training, 263 validation, and 259 testing), and the Dataset for Marine Vessel Detection from Sentinel-2 Images in the Finnish Coast (S2_FC Luhtala et al. (2024), comprising 1,295 images (906 training, 196 validation, and 193 testing). All datasets target a single object class and were split into training, validation, and test subsets for a total of 16,096 and 1,716 annotations for the pretraining and the fine tuning phases, respectively. Table 2 summarizes the partitioning details and data distribution for pretraining and fine-tuning.

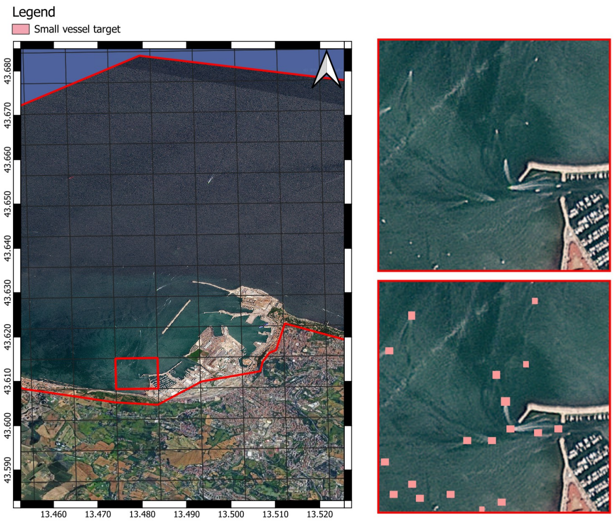


Figure 2. Tiling process applied to the Marche region case study. **Left:** Zoomed-in view of the Ancona port area with the June PlanetScope Scene as the background. **Top right:** Example of a manually selected tile. **Bottom right:** Manual annotation process performed in the Label Studio website.

Table 2. Dataset split for the pretraining and fine-tuning phases.

Pretraining			Fine-tuning		
Split	**Tiles**	**%**	**Split**	**Tiles**	**%**
Train	2942	70.00%	Train	496	70.02%
Val	630	14.99%	Val	105	14.87%
Test	630	14.99%	Test	105	14.87%
Total	**4,202**	**100%**	**Total**	**706**	**100%**

2.3 Model Architecture

The detection experiments were conducted using the YOLOv26 framework (Figure 3) Hidayatullah and Tubagus (2026), a single-stage end-to-end object detector designed to improve computational efficiency and deployment readiness in real-time applications. According to its original technical documentation, YOLOv26 introduces several architectural and optimization advances over previous YOLO implementations, including end-to-end NMS-free inference that eliminates post-processing suppression for faster and cleaner predictions; the MuSGD optimizer, a momentum-enhanced stochastic gradient descent variant designed for more stable convergence; the ProgLoss supervision, a progressive loss strategy that refines predictions across training stages; and the Small-Target-Aware Label Assignment (STAL) strategy, which improves detection accuracy by prioritizing small object matching during training. These design choices make the framework particularly suitable for the detection of small maritime targets, where dense background clutter and limited target size can reduce localization accuracy. From an architectural perspective, YOLOv26 follows a convolutional backbone combined with PAN/FPN-style multiscale feature aggregation, enabling the extraction of hierarchical features at different spatial scales. Although YOLOv26 supports multiple computer vision tasks, including object detection, instance segmentation, pose estimation, oriented detection, and classification Hidayatullah and Tubagus (2026), only the object-detection branch was considered in this study. To assess the effect of model scaling on vessel detection performance, multiple YOLOv26 variants were tested, namely the small (S), medium (M), large (L), and extra-large (X) configurations. These variants differ in terms of backbone depth, channel width, and prediction-head complexity, thus providing different trade-offs between representational capacity and computational cost. In addition, a custom pruned version of the large model, denoted as `yolo26l_cut` (Figure 3), was implemented by removing the medium- and large-scale detection heads and focusing exclusively on the P3 scale, which is particularly suited for small object detection Hidayatullah and Tubagus (2026). As illustrated in the figure, the red box highlights the pruned portion of the network. Specifically, we removed the deeper feature extraction layers corresponding to the P4 and P5 scales in both the backbone and the detection head, including the associated Conv and C3k2 blocks, as well as the aggregation modules (e.g., Concat operations) that fuse multi-scale features. By restricting the architecture to operate solely at the P3 scale, the resulting `yolo26l_cut` model reduces computational overhead while improving detection performance on small targets. This design choice is motivated by the need to enhance detection performance for small objects, which represent the primary targets in the considered coastal monitoring scenario. The adopted multi-variant setup enables a systematic analysis of the relationship between model complexity and detection performance, with the aim of identifying the most suitable YOLOv26 configuration for small-vessel detection in heterogeneous coastal imagery.

2.4 Two-stage training strategy

In the pretraining phase, models were trained on three publicly available datasets (SSDD, S2_DE, and S2_FC) to learn generalizable vessel detection features across heterogeneous data sources. During this stage, hyperparameters were selected through a systematic grid search procedure over key configuration parameters, including learning rate, batch size, and model scaling factors (e.g., architectural variants and width multipliers), with selection based on validation performance. The full set of hyperparameters explored during the pretraining phase is reported in Table 3, with the best-performing configurations highlighted in bold. Optimization is performed using the MuSGD optimizer, with the selected best hyperparameters corresponding to an initial learning rate (`lr0`) of 0.01, a final learning-rate factor (`lrf`) of 0.01, momentum of 0.937, and weight decay of 0.0005. A warm-up phase of 3 epochs was applied in all runs. Loss-related hyperparameters were kept fixed across experiments, with box, classification, and DFL loss gains set to 7.5, 0.5, and 1.5, respectively. Data augmentation was applied during pretraining and kept consistent across runs, including mosaic augmentation, horizontal flipping, HSV colour jitter, random erasing, and `RandAugment`.

In the fine-tuning phase, the pretrained models were adapted to our optical dataset to specialize in the target domain. Given

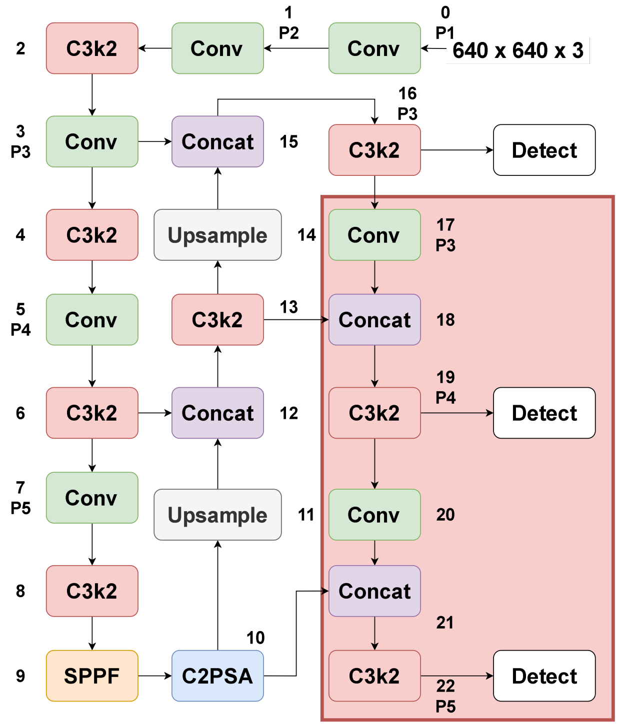


Figure 3. Pruned YOLOv26 architecture. The red box highlights the layers that have been removed.

the limited size of our dataset, a 5-fold cross-validation strategy was employed and performance was reported as the average across all folds. During this stage, a reduced learning rate was adopted and, when appropriate, the backbone was partially frozen to preserve the learned representations, while keeping the remaining training settings unchanged. No additional hyperparameter search was performed in this phase, ensuring controlled adaptation to the target domain. To investigate the influence of model scale and training configuration on vessel detection performance, multiple YOLOv26 variants were considered, including the Small, Medium, Large, and Extra-Large versions. In addition, a custom pruned configuration, denoted as `yolo26l_cut`, was derived from the Large model by removing the medium- and large-scale detection heads to assess its suitability for small-object detection. All experiments were conducted using an input resolution of 640×640 pixels and a training duration of 200 epochs. To isolate the impact of architectural scaling and training load, the main variables across runs were the model configuration and batch size, with batch sizes of 32, 64, 128, and 256 explored depending on the selected variant. Only successful runs, defined as those producing a valid `best.pt` checkpoint, were retained for comparative analysis.

All experiments were carried out on a Linux server equipped with an NVIDIA A6000 GPU (48 GB memory), within a Docker container running Ultralytics 8.4.14 on a single GPU. Evaluation was performed on the held-out test split using the best checkpoint selected according to the YOLO fitness score, defined as $\mathcal{F} = 0.1, \mathrm{mAP50} + 0.9, \mathrm{mAP50\text{-}95}$.

3. Results

The quantitative results obtained on the test set are reported in Table 4 and Table 5. Across the standard YOLOv26 variants showed stable and competitive detection performance, with

Table 3. Hyperparameter grid search space.

Hyperparameter	**values**
`lr0`	[0.001, **0.01**, 0.05]
`lrf`	[0.001, **0.01**, 0.1]
`momentum`	[0.9, **0.937**, 0.98]
`weight_decay`	[0.0001, **0.0005**, 0.001]
`warmup_epochs`	[1, **3**, 5]

mAP50 values ranging from 0.775 to 0.801 and F1-Scores between 0.756 and 0.829.

Among the standard configurations (Table 4), `YOLOv26-M` achieved the highest Precision (0.813), Recall (0.846), F1-Score (0.829), Accuracy (0.719), and mAP50-95 (0.306), confirming it as the most balanced variant in terms of detection quality and localization robustness under stricter IoU thresholds. `YOLOv26-X` obtained the highest mAP50 (0.801), yet at the cost of significantly higher computational requirements (55.635 M parameters, 193.368 GFLOPs) and the lowest inference speed (81.04 FPS). `YOLOv26-L` and `YOLOv26-X` yielded identical F1-Scores of 0.798, while the former required considerably fewer parameters (24.747 M vs. 55.635 M). `YOLOv26-S` provided the most efficient trade-off, reaching an mAP50 of 0.775 and an F1-Score of 0.756 with the fewest parameters (9.466 M), the lowest computational cost (20.522 GFLOPs), and the fastest inference (5.258 ms, 190.19 FPS).

The ablation study (Table 5) explored pruned YOLOv26 variants and lightweight YOLOv10-based configurations. Notably, `YOLOv26-L-cut` achieved the highest mAP50 across both tables (0.818) and the best Recall (0.847), while reducing the parameter count from 24.747 M to 16.388 M compared with the full `YOLOv26-L`. This suggests that architectural pruning can, in some cases, improve generalization on the adopted dataset. `YOLOv26-M-cut` showed a moderate performance decrease (F1 of 0.775, mAP50 of 0.778) but offered the lowest latency in the ablation group (5.413 ms, 184.74 FPS), making it a viable option for latency-sensitive deployment scenarios.

The YOLOv10-based light variants exhibited competitive results. `YOLOv10-M-light` reached an mAP50 of 0.802 and an F1-Score of 0.807 with only 11.234 M parameters and 51.558 GFLOPs, representing the most parameter-efficient configuration across all experiments. `YOLOv10-L-light` achieved the highest Accuracy (0.743) among all tested models, with an F1-Score of 0.806 and an mAP50 of 0.795.

In addition, the consistently lower mAP50-95 values compared with mAP50 for all configurations highlight the increased difficulty of maintaining accurate vessel localization under more restrictive overlap criteria. Considering the overall results, `YOLOv26-M` emerged as the most effective standard variant, providing the best balance between detection accuracy, retrieval capability, and localization quality, while `YOLOv26-L-cut` and `YOLOv10-M-light` stand out as compelling alternatives when model efficiency is a priority. Figure 4 provides representative qualitative examples of detection outputs for different YOLO variants on the selected test scenes.

Table 4 reports the performance of the four standard YOLOv26 compound-scaled variants on the Sentinel-2/PlanetScope test split. Overall, the results indicate that detection quality does not increase monotonically with model scale. Among the standard variants, YOLOv26-M provides the most balanced performance, whereas larger configurations incur higher computational costs without yielding consistent improvements across the evaluated criteria. This pattern suggests that, for the present small-vessel detection task, increasing model capacity alone is not sufficient to guarantee better overall performance.

An additional observation concerns the YOLOv26-L variant, which does not improve upon the overall performance of YOLOv26-M despite its higher model capacity. This result suggests that, in the present small-vessel detection setting, increasing architectural complexity does not necessarily translate into better generalization. A possible explanation is that larger configurations may be less well matched to the scale of the available training data and to the small-object characteristics of the task. Finally, from an efficiency perspective, YOLOv26-S appears to be the most suitable option for latency-constrained scenarios, as it combines the lowest computational burden with the fastest inference among the standard variants, albeit at the expense of lower overall detection performance than YOLOv26-M. By contrast, YOLOv26-X increases computational cost substantially and does not provide a commensurate improvement in overall performance, indicating diminishing returns from further scaling in the present experimental setting.

3.1 Ablation study

We investigated the impact of architectural simplifications and backbone design choices on vessel detection performance. The standard YOLOv26 head produced predictions at three scales, namely P3/8 (small objects), P4/16 (medium), and P5/32 (large). In the proposed `cut` variants (L-cut, M-cut) the P4 and P5 detection branches were entirely removed, including their associated upsampling, concatenation and C3k2 fusion layers. The backbone (composed of C3k2 blocks, SPPF, and C2PSA modules) was kept unchanged, while the detection head was reduced to a single output at the P3/8 scale. This modification reduced model complexity, leading to a parameter reduction of approximately 35% (e.g., 24.7M to 16.4M for the L variant) and a decrease of about 17% in GFLOPs, while preserving high-resolution features suitable for small-object detection.

In parallel, we explored `light` variants (L-light, M-light) based on a YOLOv10-inspired backbone, originally introduced in our previous work Galdelli et al. (2025). In these models, the backbone differs from the YOLOv26, using `C2f`/`C2fCIB` blocks instead of `C3k2`, `SCDown` in place of standard strided convolutions, and `PSA` instead of `C2PSA`. Similarly to the cut configurations, these variants adopted a single-scale detection head operating exclusively at P3. Detection was performed using the NMS-free `v10Detect` head. Among these configurations, the M-light variant ($d = 0.67$, $w = 0.75$, corresponding to depth and width scaling factors, respectively) achieved the lowest model complexity (11.2M parameters and 51.6 GFLOPs), offering the highest inference efficiency at the cost of a moderate reduction in accuracy.

4. Discussions

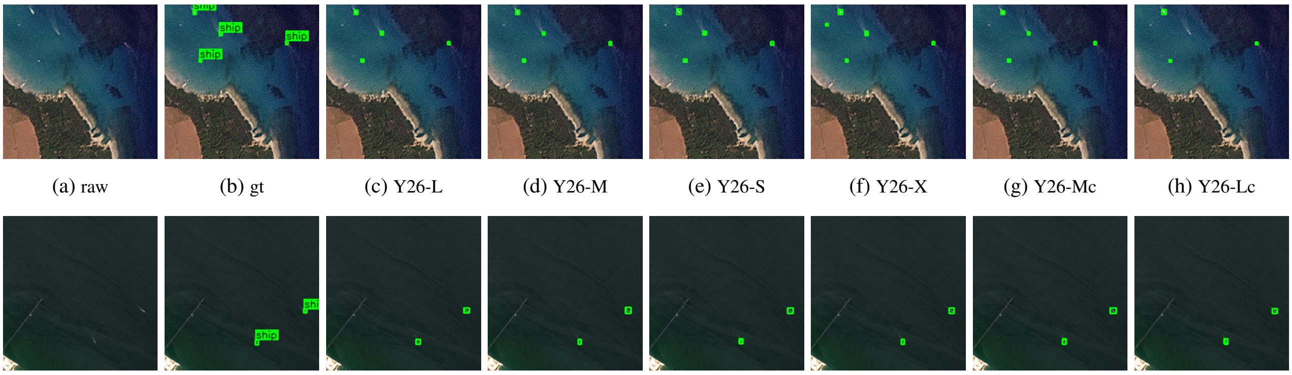
The experiments conducted in this study provided empirical evidence on the behaviour of different deep learning architectures for small-vessel detection in heterogeneous coastal optical imagery. Consistent with the broader literature on small-object detection in remote sensing, the results suggest that target size, image resolution, and detector design jointly influence performance, which may reduce the benefits of simply increasing model

Table 4. Performance of standard YOLOv26 variants S/M/L/X

Model	Precision	Recall	F1-Score	Accuracy	mAP50	mAP50-95	Latency (ms)	GFLOPs	FPS	Params (M)
YOLOv26-S	0.764	0.748	0.756	0.685	0.775	0.293	**5.258**	**20.522**	**190.19**	**9.466**
YOLOv26-M	**0.813**	**0.846**	**0.829**	**0.719**	0.785	**0.306**	5.946	67.848	168.18	20.350
YOLOv26-L	0.802	0.795	0.798	0.649	0.796	0.303	8.628	86.095	115.90	24.747
YOLOv26-X	0.782	0.814	0.798	0.709	**0.801**	0.282	12.340	193.368	81.04	55.635

Table 5. Ablation study results

Model	Precision	Recall	F1-Score	Accuracy	mAP50	mAP50-95	Latency (ms)	GFLOPs	FPS	Params (M)
YOLOv26-L	0.802	0.795	0.798	0.649	0.796	0.303	8.154	86.095	122.64	24.747
YOLOv26-M	**0.813**	0.846	**0.829**	0.719	0.785	**0.306**	6.350	67.848	157.48	20.350
YOLOv26-L-cut	0.783	**0.847**	0.814	0.738	**0.818**	0.286	8.080	71.053	123.76	16.388
YOLOv26-M-cut	0.781	0.769	0.775	0.686	0.778	0.282	**5.413**	55.532	**184.74**	12.845
YOLOv10-L-light	0.802	0.811	0.806	**0.743**	0.795	0.283	8.111	116.332	123.29	21.283
YOLOv10-M-light	0.806	0.808	0.807	0.724	0.802	0.289	5.691	**51.558**	175.72	**11.234**



(a) raw (b) gt (c) Y26-L (d) Y26-M (e) Y26-S (f) Y26-X (g) Y26-Mc (h) Y26-Lc

(i) raw (j) gt (k) Y26-L (l) Y26-M (m) Y26-S (n) Y26-X (o) Y26-Mc (p) Y26-Lc

Figure 4. Comparison between raw images, ground truth, and different YOLO variants on Brindisi (top row, PlanetScope) and Ancona (bottom row, Sentinel-2), showing robustness across heterogeneous data.

complexity Wang et al. (2023). The comparison among the standard YOLOv26 variants indicated that increasing model capacity does not lead to proportional performance gains in the evaluated setting. As observed in our comparative analysis, scaling up the model capacity from the medium (YOLOv26-M) to the extra-large variant (YOLOv26-X) increased computational costs and GFLOPs without yielding proportional gains in detection quality. Within the limits of the present study, this suggests that a medium-capacity configuration can provide an effective trade-off between model complexity and the challenges posed by small-target detection in coastal satellite imagery. Conversely, larger models increased computational cost without providing proportional gains in performance. Although YOLOv26-X had the highest parameter count and GFLOPs, it did not improve the overall performance balance achieved by YOLOv26-M. Overall, the results suggested that, in this single-class small-target detection setting, increasing model capacity alone does not necessarily improve generalization. This interpretation was consistent with recent reviews showing that small-object detection often remains constrained by limited target footprint, feature loss through downsampling, and localization instability, even when more complex detectors are adopted Yu et al. (2026). Another relevant interpretative element is the systematic discrepancy observed between mAP50 and mAP50-95. This gap suggests that the evaluated architectures were more effective at detecting the presence of the target than at achieving highly precise spatial localization under stricter overlap thresholds. Since mAP50-95 aggregates performance over increasing IoU thresholds, this effect is consistent with observations in the small-object detection literature, where IoU-based localization metrics are highly sensitive to minor pixel-level spatial deviations when the target footprint is very limited Jeune and Mokraoui (2023). Consequently, the performance degradation observed at higher IoU thresholds should not be attributed solely to limitations of the detector itself, but also to geometric constraints intrinsic to IoU-based evaluation for very small targets. Overall, these findings supported a cautious interpretation, namely that in SSF vessel detection over heterogeneous coastal optical imagery, task-oriented architectural design choices may be more informative than simply increasing detector capacity. This conclusion was consistent with recent remote-sensing studies showing that dedicated small-object detection strategies, including scale-aware feature handling and lightweight architectural optimization, can be more effective than brute-force scaling alone Yao et al. (2025).

5. Conclusions and future work

In this study we investigated the challenging task of detecting SSF vessels in heterogeneous coastal environments using a multi source optical dataset, and combining extensive pre-training on public datasets with domain specific fine-tuning. The comprehensive comparative analysis of YOLOv26 architectures demonstrated that, for small-target detection, increasing model capacity does not inherently translate into proportional performance gains. Within the standard configurations, YOLOv26-M provided the most effective balance between representational power and operational efficiency. Furthermore, the ablation study revealed that heavy architectures suffer from diminishing returns, whereas task-oriented simplifications (i.e., pruning the deeper P4 and P5 detection branches to focus exclusively on the high-resolution P3 scale) yielded highly competitive results with a substantially reduced computational footprint. This confirms that small-vessel detection is primarily bottlenecked by target scale, downsampling feature loss and localization hypersensitivity under strict IoU thresholds, rather than by a deficit in semantic complexity. Consequently, lightweight and scale-aware architectures emerge as the most viable solutions for resource-constrained, near real-time maritime surveillance. While the current framework provides a robust benchmark for single-class vessel detection in optical imagery, several avenues remain open for future development. From an architectural standpoint, the promising performance of the pruned variants warrants further investigation into advanced model compression techniques, such as knowledge distillation and quantization, to facilitate deployment on edge devices or onboard satellite processing units. From a data perspective, the observed overfitting of larger models (e.g., YOLOv26-L) underscores that the current dataset limits the exploitation of higher-capacity networks.

Future research must prioritize the systematic expansion of the dataset across diverse geographic, seasonal and atmospheric conditions. A larger and more heterogeneous corpus is essential not only to improve the external validity and generalization of the models but also to enable the transition from single-class detection to multi-class vessel classification. Finally, to establish a fully comprehensive coastal monitoring framework, future efforts should extend the methodological workflow to include explicit experimental validation on SAR data, alongside exploring multi-temporal analyses for trajectory inference and anomaly detection.

6. Acknowledgement

The research activity is funded by the Horizon Europe framework programme under Grant Agreement No. 101190021, within the project AI4COPSEC (https://ai4copsec.eu/).

References

Anderson, D., Belcineanu, A., Tzvetkova, M., 2025. AI for Border Management and Customs Controls.

Christodoulou, A., Echebarria Fernández, J., 2021. Maritime governance and international maritime organization instruments focused on sustainability in the light of united nations' sustainable development goals. *Sustainability in the maritime domain: towards ocean governance and beyond*, Springer, 415–461.

European Parliament and Council of the European Union, 2013. Regulation (EU) No 1380/2013 on the Common Fisheries Policy. *Official Journal of the European Union*, L 354, 22–61. https://eur-lex.europa.eu/eli/reg/2013/1380/oj.

European Parliament and Council of the European Union, 2023. Regulation (EU) 2023/2842 of the European Parliament and of the Council of 22 November 2023 amending Council Regulation (EC) No 1224/2009. *Official Journal of the European Union*, L. https://eur-lex.europa.eu/eli/reg/2023/2842/oj.

European Space Agency, 2026a. S2 mission. https://sentiwiki.copernicus.eu/web/s2-mission. Accessed: 15 March 2025.

European Space Agency, 2026b. Sentinel-1: Facts and figures. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Facts_and_figures. Accessed: 18 March 2026.

European Space Agency, 2026c. Sentinel-2: Facts and figures. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Facts_and_figures. Accessed: 18 March 2026.

Galdelli, A., Mancini, A., Ferrà, C., Tassetti, A. N., 2021. A synergic integration of AIS data and SAR imagery to monitor fisheries and detect suspicious activities. *Sensors*, 21(8), 2756.

Galdelli, A., Mancini, A., Ferrà, C., Tassetti, A. N. et al., 2020. Integrating ais and sar to monitor fisheries: A pilot study in the adriatic sea. *Proceedings of the 2020 IMEKO TC-19 International Workshop on Metrology for the Sea, Naples, Italy*, 5–7.

Galdelli, A., Narang, G., Pietrini, R., Zazzarini, M., Fiorani, A., Tassetti, A. N., 2025. Multimodal AI-enhanced ship detection for mapping fishing vessels and informing on suspicious activities. *Pattern Recognition Letters*, 191, 15–22.

Guan, T., Chang, S., Wang, C., Jia, X., 2025. SAR Small Ship Detection Based on Enhanced YOLO Network. *Remote Sensing*, 17(5), 839.

Hidayatullah, P., Tubagus, R., 2026. YOLO26: A Comprehensive Architecture Overview and Key Improvements. *arXiv preprint*.

Jeune, P. L., Mokraoui, A., 2023. Rethinking intersection over union for small object detection in few-shot regime. *arXiv preprint arXiv:2307.09562*.

La, T.-V., Pham, M.-T., 2024. Insight into the collocation of multi-source satellite imagery for multi-scale vessel detection. *arXiv preprint arXiv:2403.13698*.

Lattanzi, P., Pulcinella, J., Battaglia, P., Di Cintio, A., Ferrà, C., Di Franco, A., Tassetti, A. N., 2024. Bridging the gap in fishing effort mapping: a spatially-explicit fisheries dataset for Campanian MPAs, Italy. *Scientific Data*, 11(1), 54.

Li, F., Yu, K., Yuan, C., Tian, Y., Yang, G., Yin, K., Li, Y., 2025. Dark Ship Detection via Optical and SAR Collaboration: An Improved Multi-Feature Association Method Between Remote Sensing Images and AIS Data. *Remote Sensing*, 1–26.

Luhtala, J., Virtanen, J., Silvennoinen, S., 2024. A dataset for marine vessel detection from sentinel-2 images in the finnish coast. https://doi.org/10.5281/zenodo.10046342. Accessed: 15 November 2025.

Orofino, S., McDonald, G., Mayorga, J., Costello, C., Bradley, D., 2023. Opportunities and challenges for improving fisheries management through greater transparency in vessel tracking. *ICES Journal of Marine Science*, 80(4), 675–689.

Planet Labs PBC, 2024. PlanetScope Product Specifications. https://assets.planet.com/docs/Planet_PSScene_Imagery_Product_Spec_letter_screen.pdf. Accessed: 18 March 2026.

Planet Labs PBC, 2025. Planetscope scene product specification. https://planet.widen.net/s/vswqcrm9zw/planet-userdocumentation-psscene-productspec. Accessed: 15 January 2025.

Roboflow Universe, 2023. Sentinel-2 Ship Detection Dataset. `https://universe.roboflow.com/sentinel2/sentinel-2-ship_detection`. Dataset created by sentinel2. Accessed: 15 March 2025.

Shepperson, J. L., Hintzen, N. T., Szostek, C. L., Bell, E., Murray, L. G., Kaiser, M. J., 2018. A comparison of VMS and AIS data: the effect of data coverage and vessel position recording frequency on estimates of fishing footprints. *ICES Journal of Marine Science*, 75(3), 988–998.

Sumaila, U. R., Zeller, D., Hood, L., Palomares, M. L. D., Li, Y., Pauly, D., 2020. Illicit trade in marine fish catch and its effects on ecosystems and people worldwide. *Science Advances*, 6(9), eaaz3801.

Tkachenko, M., Malyuk, M., Holmanyuk, A., Liubimov, N., 2022. Label studio: Data labeling software. `https://github.com/heartexlabs/label-studio`. Open-source software, Accessed: 15 September 2025.

Wang, X., Wang, A., Yi, J., Song, Y., Chehri, A., 2023. Small Object Detection Based on Deep Learning for Remote Sensing: A Comprehensive Review. *Remote Sensing*, 15(13).

Wang, Y., Wang, C., Zhang, H., Dong, Y., Wei, S., 2019. A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds. *Remote Sensing*, 11(7), 765.

Yao, B., Zhang, C., Meng, Q., Sun, X., Hu, X., Wang, L., Li, X., 2025. SRM-YOLO for Small Object Detection in Remote Sensing Images. *Remote Sensing*, 17(12).

Yu, H., Liu, J., Lin, M., 2026. A Comprehensive Literature Review on YOLO-Based Small Object Detection: Methods, Challenges, and Future Trends. *Computers, Materials and Continua*, 87(1).

Zhang, T., Zhang, X., Li, J., Xu, X., Wang, B., Zhan, X., Xu, Y., Ke, X., Zeng, T., Su, H., Ahmad, I., Pan, D., Liu, C., Zhou, Y., Shi, J., Wei, S., 2021a. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sensing*, 13(18), 3690.

Zhang, T., Zhang, X., Li, J., Xu, X., Wang, B., Zhan, X., Xu, Y., Ke, X., Zeng, T., Su, H., Ahmad, I., Pan, D., Liu, C., Zhou, Y., Shi, J., Wei, S., 2021b. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sensing*, 13(18).