

# Detecting two-pixels long boats on satellite images



Author: Luka LAVAL  
Supervisors: David MOUILLOT, Pauline VIGUIER  
UMR MARBEC - 08/2025

I want to thank the whole MARBEC team, especially the wonderful Building 15 crew: Pauline, Auguste and Gaetan for their precious help and their terrible office jokes; Martin, Lily and Jeanne for the much-needed table tennis breaks; without forgetting Isabelle, Sarah and Siyu; Lucas aka "the cool guy from another lab"; our most recent member, Taha; and of course David, without whom this experience would not have been possible.

Thank you

The code of the project, models weights, and datasets are accessible on GitHub:  
<https://github.com/LUKALAVAL/boat-detection>

For any other information please contact me by email: [lukalavalferk@gmail.com](mailto:lukalavalferk@gmail.com)

I used GitHub Copilot to assist me during coding to produce cleaner implementations. Additionally, ChatGPT and QuillBot were used to help structure paragraphs and refine the English phrasing, particularly in the introduction.

# Table of contents

<b>1 Abstract.....</b>	<b>5</b>
<b>2 Context.....</b>	<b>5</b>
<b>3 Introduction and state of the art.....</b>	<b>5</b>
<b>4 Methods.....</b>	<b>8</b>
4.1 Dataset construction.....	8
4.2 Model.....	11
4.3 Inference.....	11
4.4 Filtering.....	13
<b>5 Results.....</b>	<b>15</b>
5.1 Model evaluation.....	15
5.2 Database evaluation.....	17
<b>6 Limitations and discussion.....</b>	<b>20</b>
<b>7 Conclusion.....</b>	<b>21</b>
<b>8 References.....</b>	<b>22</b>
<b>9 Appendix.....</b>	<b>24</b>

# 1 Abstract

Monitoring small vessel activity is critical for understanding human pressure on marine and coastal ecosystems, yet such vessels often lack AIS (Automatic Identification System) transponders and typically require costly, very high-resolution satellite imagery for reliable detection that may be sparse in time and space. We developed a boat detection model that identifies vessels under 6 meters in length from PlanetScope 3m/px satellite imagery by training on degraded Pléiades high-resolution images. This enables scalable mapping of small-boat activity across broad spatial and temporal extents, providing valuable proxies for human activity in ecological studies, such as species distribution modeling.

## 2 Context

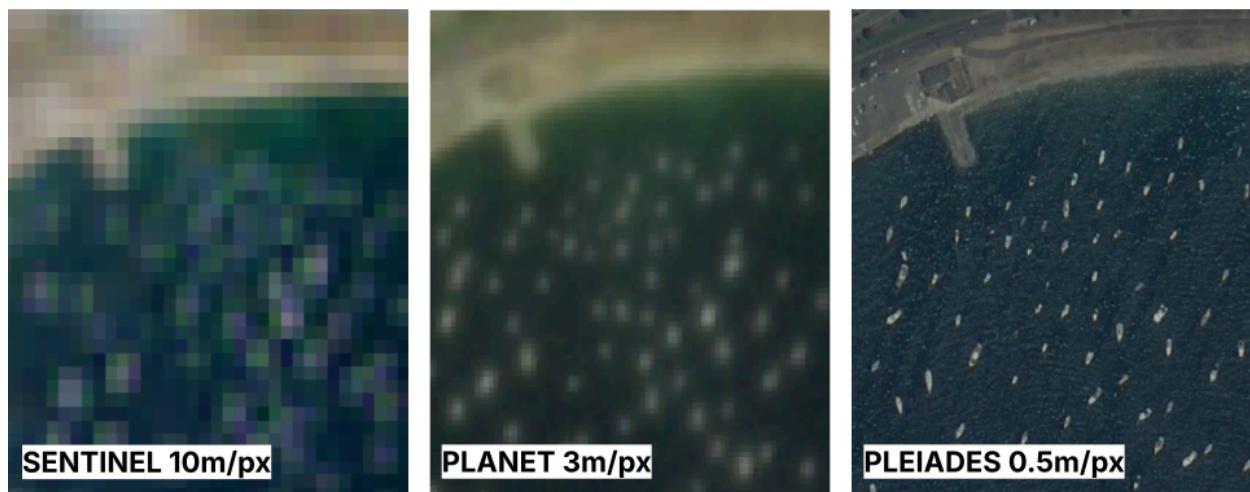
This report was written as part of my Master's internship in Computer Science (M2 IMA) at Sorbonne University, conducted within the joint research unit MARBEC (MARine Biodiversity, Exploitation and Conservation), under the supervision of David Mouillot and Pauline Viguier. MARBEC is one of the largest French research units dedicated to marine sciences. I was well supported by the team of engineers and fellow interns, and also had the opportunity to regularly interact with marine researchers, which provided necessary insights for real-world applications of my work.

## 3 Introduction and state of the art

Quantifying maritime traffic is an evolving task. The arrival of the AIS (Automatic Identification System) in the early 2000s transformed vessel tracking: based on VHF communication, AIS provide vessel coordinates, speed, heading and a unique identifier (MMSI) at regular intervals (Le Berre et al., 2024). Since 2007, the International Maritime Organization has required AIS on larger commercial vessels (over 300 tons) or carrying more than 165 passengers. In Europe, since 2014, the obligation has extended to fishing vessels over 15 meters in length. AIS data are captured by terrestrial receivers, effective up to 350 nautical miles (around 650km) from the coast, by satellites covering remote areas and by nearby equipped boats (AIS (*Automatic Identification System*) Overview, 2021). AIS data can be purchased on distribution platforms at various temporal resolutions (down to two minutes interval) and is used to monitor marine traffic (Datalastic, n.d.; MarineTraffic, n.d.). Nevertheless, AIS coverage is incomplete, as some vessels are either intentionally or unintentionally unequipped (Paolo et al., 2024). To fill these gaps, AIS is often paired with satellite remote sensing; but the challenge remains: the boats still need to be detected in the images.

Observation satellites use either passive sensors (detecting waves emitted by the Earth) or active sensors (radar waves emitted and then reflected eg. SAR). Boat detection from Synthetic Aperture Radar (SAR) images is widely used because Sentinel-1 SAR images are free, they have a large spatial coverage, SAR works for both day and night observations, and the radar waves can penetrate clouds (Ager, 2013). It can even detect boats smaller than the sensor resolution (Gower & Skey, 2000). However, Sentinel-1 has a rather low temporal resolution (about 6 days revisit frequency), and the radar backscatter is affected by the material, making non-metal boats more difficult to detect (Pelich et al., 2015; *Sentinel-1 Mission*, n.d.). In addition to radar, optical sensors are also used for vessel detection. These

passive instruments capture reflected sunlight that can reveal not only the presence of ships but also their shapes and dimensions. Their main limitation is dependence on daylight and clear weather conditions (Li et al., 2021). There is a large range of resolutions accessible from roughly 10m/px with Sentinel-2 down to 0.5m/px with Pléiades (*Pléiades Full Archive and Tasking*, n.d.; *Sentinel-2 Mission*, n.d.). Optical satellite images can be costly in exchange for a higher spatial or temporal resolution. Generally speaking, objects occupying about 3 pixels can be distinguished (Bannister & Neyland, 2015). We used 3m/px PlanetScope optical satellite imagery for its large spatial and temporal coverage (about 1 visit per day since 2016). The satellite image resolution is just good enough to visually discriminate small boats from each other, but too coarse to recognize them formally. Overall, Planet is a good compromise at a moderate price (*PlanetScope*, 2025). It is important to mention that artificial image resolution enhancement techniques cannot be directly applied to boats as they are moving targets, unlike buildings (Meares, 2024).



*Figure 1. Small boats close to the coast from 3 optical sensors. On the left, the boats merge into clusters of bright pixels (Sentinel 2). In the center, the boats are distinguishable but are only white spots (Planet). On the right, the boats are distinguishable and recognizable (Pléiades).*

Modern learning-based techniques and conventional image-processing procedures are the two main categories of boat detection methods. Adaptive threshold based methods consist of defining a segmentation threshold based on local or global statistics to extract the boats from the image. Simple segmentation, edge detection, or morphological filtering can also be used to interpret optical data. All these methods can detect larger vessels, but they frequently struggle when the targets are close to the sensor's resolution limit (Kanjir et al., 2018). Object detection frameworks using convolutional neural networks (CNN) (e.g., Faster R-CNN, YOLO) have been applied to remote sensing imagery for boat detection (Moujahid et al., 2023). Additional strategies are usually implemented to improve the detections on smaller and densely packed objects such as feature pyramid network (FPN) or anchor-free based implementation. FPN-based strategies combine high-resolution shallow features with high-level semantic context, enabling better performances on small boats (Liu et al., 2023). Anchor-free detection benefits small-boat detection by removing dependence on predefined anchor sizes and locations, allowing the model more flexibility (Wen et al., 2023). The latest YOLO models use an FPN-based method with an anchor free implementation that facilitates object detection in cluttered scenes (Khanam & Hussain, 2024). Bounding boxes are used in most cases, they can either be aligned or oriented. Usually, oriented bounding boxes (obb)

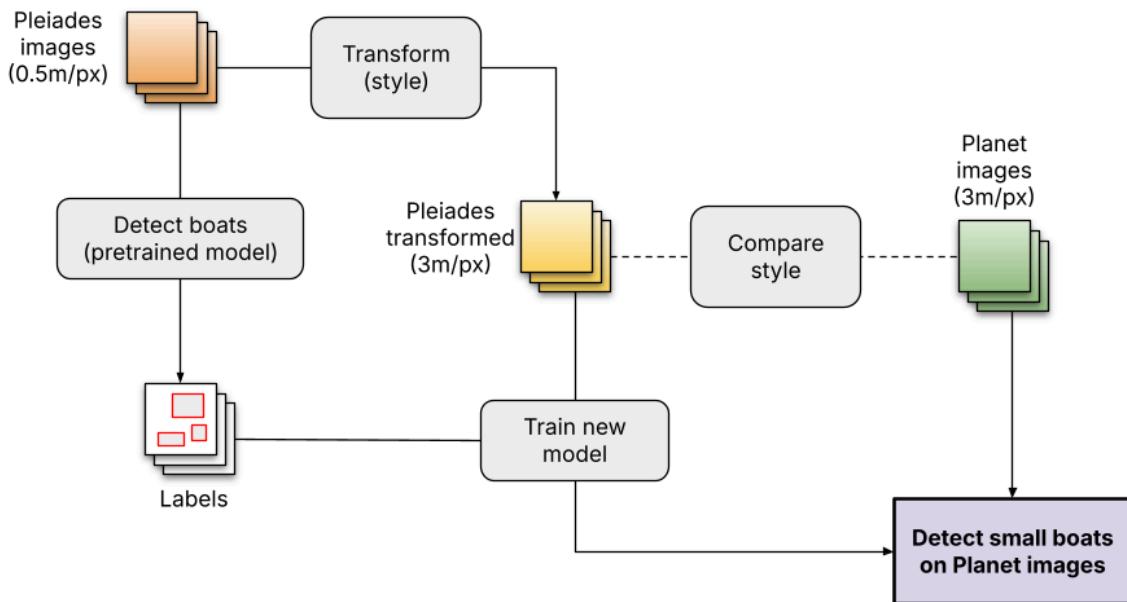
are preferred for boat detection in satellite imagery because ships can appear at arbitrary angles, and axis-aligned boxes often enclose excessive background water or overlap with neighboring vessels in crowded regions (Wen et al., 2023). Additionally, obb preserves relevant features such as vessel length and aspect-ratio that can be used for filtering. We have also briefly explored a point detection model pre-trained for white spots detection on microscope imagery (see section 8 Limitations and discussion) (Niu et al., 2024).

A large variety of datasets are available to train boat detection models such as tile classification based datasets, or oriented bounding boxes with multiple image modalities (Faudi & Martin, 2018; Gallego et al., 2018; Xia et al., 2018; Robert Hammell, n.d.). As the smaller boats are very difficult to formally identify on Planet images, we could not find a Planet dataset with "real" ground truth.

The goal of this project is to create a boat presence database from 2018 to 2023 around 1992 environmental DNA (eDNA) locations in the mediterranean sea. The final database is intended to be used to construct a boat density layer for a species distribution model (SDM).

## 4 Methods

We propose a method to train a boat detection model for Planet images (3m/px). This resolution is just good enough to distinguish small vessels individually but too coarse to ensure the white spots actually correspond to a boat and not sea foam or a piece of land. Actual ground truth for small boats cannot be derived directly from a Planet image. Our method consists in training the model on downsampled high resolution images (from 0.5m/px downsampled to 3m/px). We use the high resolution images to create the ground truth labels and make them correspond to the downsampled images for training. Then we use the trained model to detect boats directly on the lower resolution Planet images. We chose to work with Ultralytics' YOLOv11-obb model for the ease of implementation and flexibility of the framework. This section details the boat detection pipeline from the dataset construction to the predictions filtering.



*Figure 2. Pipeline for training an object detection model on a target image modality via style transfer*

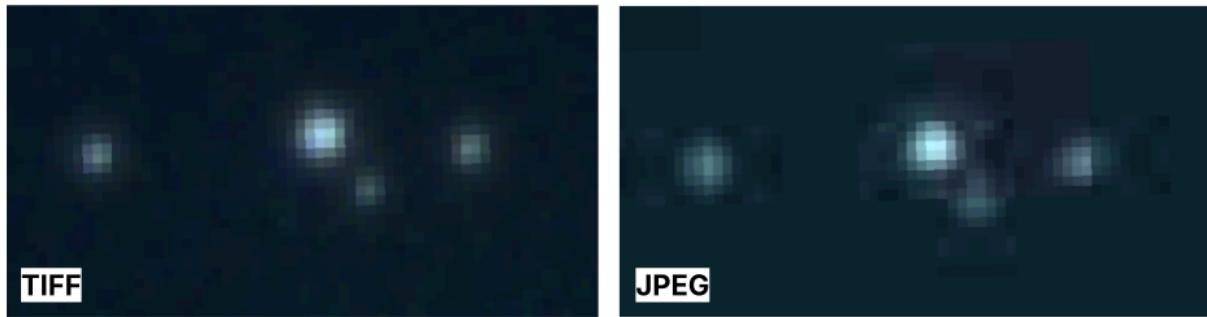
### 4.1 Dataset construction

#### Image tiles

To build the dataset, we have downloaded Pléiades satellite images available on Dinamis. We selected images containing sea, most of them are close to the french mediterranean coasts. We paid special attention to avoid overlap between the images when possible as well as favoring large images with boat presence to reduce the workload during the process. Although we did not consider seasonality, we made sure to select images containing a few boats as well.

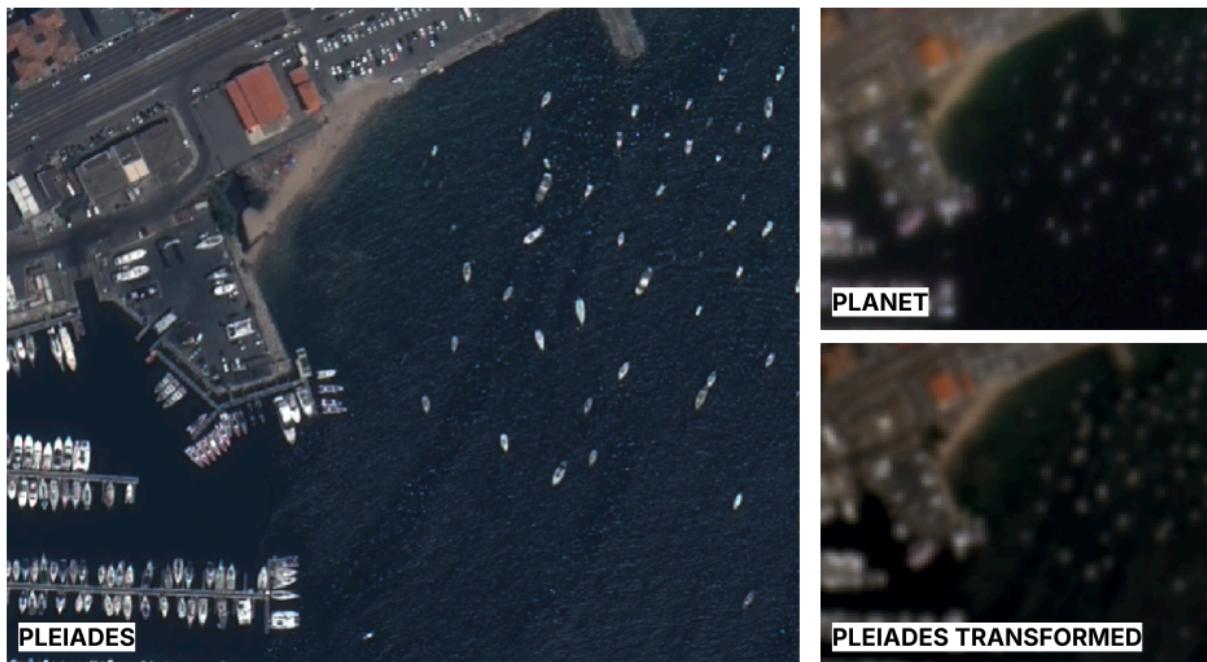
The four image layers (red, green, blue, panchromatic) were pansharpened using GDAL with lanczos resampling, and converted to 8 bits .tiff format. The pansharpening process is a

technique to enhance the resolution of the multi-spectral layers by fusing it to the higher resolution panchromatic layer. We end up with images close to 0.5m/px resolution. We have decided to work with .tiff images instead of .jpeg so as to keep smaller details on tiny boats apparent. The .jpeg compression can hide details and create artefacts on smaller objects.



*Figure 3. Four small vessels on a Planet image saved in .TIFF and .JPEG format with exaggerated compression artefacts. Those artefacts may appear when using jpeg compression on big image files.*

Each image is then transformed to match Planet imagery style. The transformation process consists in downsampling the image resolution to 3m/px using bilinear interpolation and adding a Gaussian blur with a variance of 1. Although Pléiades and Planet satellites have different capturing instruments, this simple method achieves visually convincing results as shown in the figure below. The limitations of this procedure are detailed in section 6 Discussion. The final step consists in slicing the transformed images into tiles of 512×512 pixels. The tiles without any color variance are removed. As they are outside of the image they do not show any added value and would just increase the training time.



*Figure 4. Pléiades pansharpened (0.5m/px), Planet (3m/px) and Périades after transformation function (3m/px). There are a few hours interval between Pléiades and Planet acquisitions*

## Labeling

So as to reduce the labeling cost, we applied a detection model on the high resolution Pléiades images to detect the boats. We used Ultralytics' yolo1x-obb model trained on DOTA v1 dataset with SAHI (Slicing Aided Hyper Inference) method to perform the inference on the large satellite images with a window size of  $1024 \times 1024$  pixels, an overlap of 0.1 and confidence threshold of 0.3. SAHI eases the inference process by slicing the larger image into overlapping tiles and merging the predictions accordingly (Akyon et al., 2022). The annotations generated were then georeferenced and manually verified and modified on QGIS: missing oriented bounding boxes were added, false detections were removed, and moving boats with their boat wake were reclassified as such. The smaller boats docked in harbors were also removed as they are usually not distinguishable at 3m/px. The dataset has two classes of objects labeled as "boat" and "wake". The "wake" class consists of the smaller moving boats when the wake is clearly visible on the image, the bounding box contains the visible wake. The "boat" class consists of all the remaining boats, mostly stationary boats. We used these two classes to differentiate boxes for which we can derive the size of the boat (class "boat") from boxes for which the size does not represent the size of the boat (class "wake") after inference. It also appeared that the optimal confidence thresholds of the two classes are far apart, which we wouldn't have noticed if there was a single class. This splitting was also used in *Monitoring visitation at North Carolina artificial reef sites using high spatiotemporal resolution PlanetScope imagery* (Ricci & Bohnenstiehl, 2022). Finally the label files are sliced to match the previously mentioned tiles.

A first training was launched and showed poor results, it was understood that the boxes were too small for the model to properly train on. Many boxes of the class "boat" contained less than three pixels on the downscaled images. To counter this effect all the boxes were enlarged by 6 meters (2 pixels at 3m/px) in all four directions.

## Train, validation and test sets

The final dataset is split randomly in train, validation and test sets. The train set fits the model, the validation set validates hyperparameters, the test set evaluates the model. Note that the dataset is relatively imbalanced as the "wake" class is rarer than the "boat" class in the real case scenario. We also observed a strong imbalance in the vessel sizes, with the large majority of vessels measuring less than 20 meters. To mitigate this, we added Planet images near the Strait of Gibraltar and Gulf of Suez containing container ships only to the dataset. Despite these additions, the dataset remains noticeably imbalanced.

Split	Ratio	Total tiles	Tiles with at least one object	Total "boat" class	Total "wake" class
train	0.8	5601	1086	6693	543
validation	0.1	700	116	698	90
test	0.1	701	151	1217	84

Table 1. Train, validation and test split of the dataset. There is a class imbalance in favor of the "boat" class, as "wake" are rarely captured

## Planet dataset

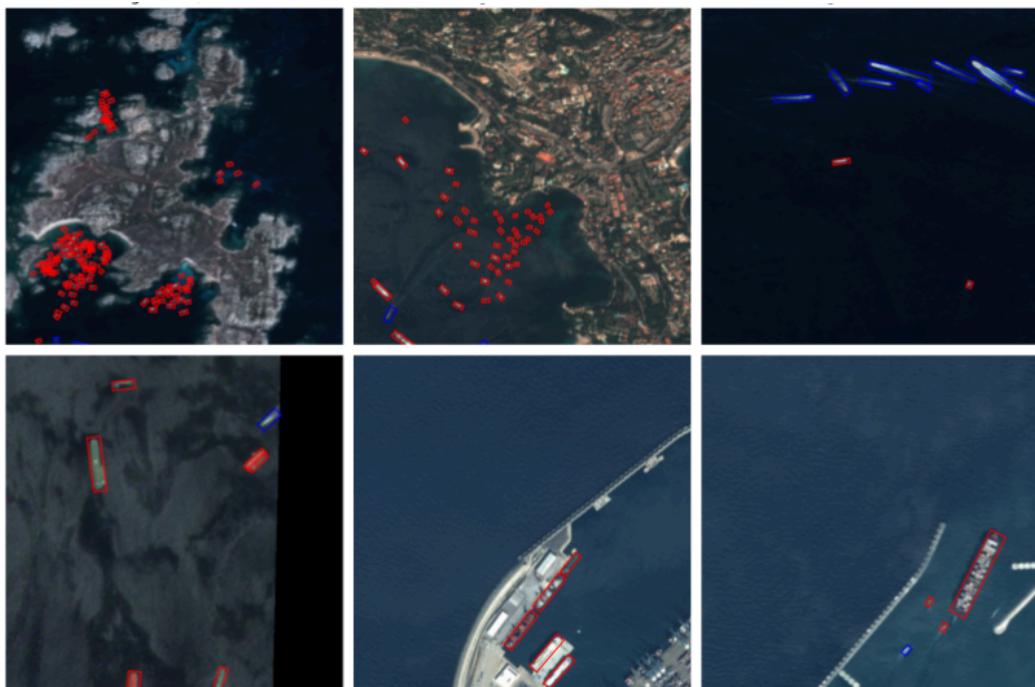
A separate dataset consisting exclusively of Planet images was also constructed. This dataset was manually labeled with a naïve approach, where small white spots were labeled as

boats whenever they were deemed likely to correspond to actual vessels. The images were also selected individually to have a variety of situations and boats, while avoiding overlapping between the images. This dataset was used to train a “naïve” model we called PLA exclusively on Planet images. PLA is later used for comparison (see section 5.1 Model evaluation).

## 4.2 Model

### Training

The oriented bounding box model was trained using Ultralytics’ framework over 256 epochs (about 4 hours) after a plateau was detected with a patience of 100 epochs. We used the default yolo11m-obb weights pretrained on DOTA v1 dataset. Our dataset was augmented for the training within the framework using hue, saturation and value transformations as well as image rotation, up-down and left-right flips, and mosaics. The training process relied on Ultralytics’ automated optimization strategy, where the framework selects the most suitable optimizer and learning rate scheduler for convergence. The composite loss function used consists of 3 main components: box regression loss, classification loss and Objectness Loss. The box regression loss measures the overlap between the predicted oriented box and the ground truth (taking into consideration scale and rotation); the classification loss optimizes the multi-class predictions using binary cross-entropy; Objectness Loss ensures the model assigns correct probabilities of object presence.

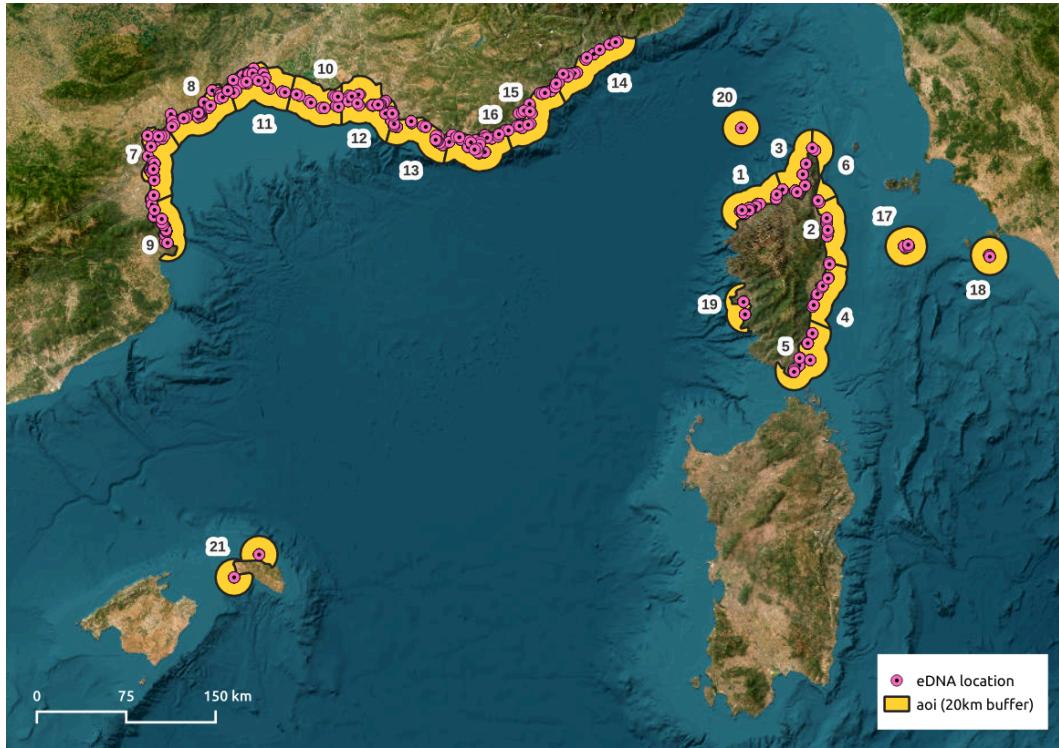


*Figure 5. Train set sample tiles annotated with “boat” class in red and “wake” class in blue*

## 4.3 Inference

This project was originally designed to add a boat density layer to an SDM (Species Distribution Model) around 1992 eDNA locations from 2018 to 2023 across the Mediterranean

sea. A 20km buffer around each location was added, from which we removed most of the land part. The area was then sliced into 21 areas of interest.



*Figure 6. Map of points of interest with 20km buffer zone splitted into 21 areas of interest*

The images for inference were automatically downloaded using Planet's API. We have decided to create composite images by day. Image composite consists in stitching images together in a specific order by removing the overlaps. This approach maximizes the downloaded area each day while minimizing the overlaps resulting in minimizing the costs. The composite order was set according to the cloud coverage on each image, favoring images with low cloud coverage first.

Instead of downloading every day, we have decided to select one day in the first five days of the week and one day during the week-end for each week from 2018 to 2023. This reduces the amount of composite images downloaded by 3.5. We tried to maximize the weighted score to select the best day:

$$0.7 \times overlap\_aoi + 0.3 \times (1 - cloud\_cover)$$

where *overlap\_aoi* corresponds to the overlap of the composite image with the area of interest (0 to 1), and *cloud\_cover* is a score based on the relative contribution (overlap) of the images in the composite image.

The inference was parallelized on 3 GPUs and took about 30 hours to complete on the 11995 composite images. We used SAHI with a window size of 512×512 pixels and an overlap of 0.1. The confidence threshold was set to 0.001 to filter the predictions during the post-processing step (see section 4.4 Filtering).

## 4.4 Filtering

To maximize the predictions accuracy, we developed a set of filters to apply on the raw predictions. In order to ease the filtering process, the predicted boxes are stored in a reduced version considering the date of the detection, its center coordinates, length, width, aspect ratio, orientation, and class.

### Cloud mask

Each composite image has a corresponding binary raster cloud mask. Although the model was trained with cloud images in the dataset, we removed each detection that has its center point on the cloud mask. This conservative approach reduces ambiguous detections.

### Land mask

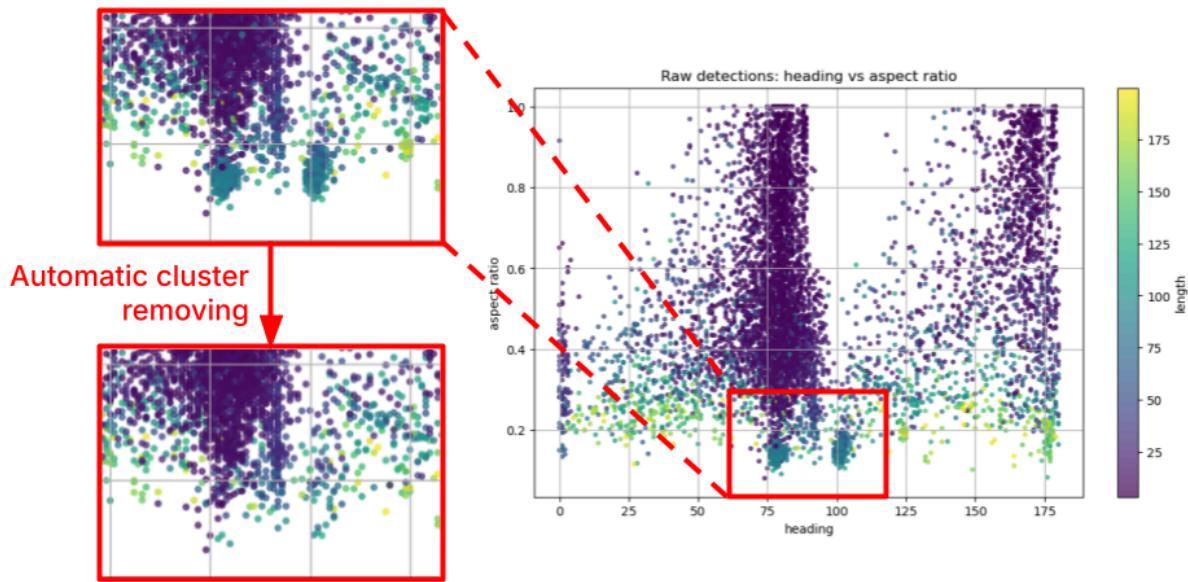
In a similar way, we removed all the detections on land although the model was trained with land images. The only difference relies on the nature of the data, while the cloud mask is a raster, the land mask is a georeferenced polygon constructed using multiple OSM (OpenStreetMap) elements (OpenStreetMap, 2017). Although the polygon is quite precise, pieces of land are missing.

Small boats can be quite close to the coast while bigger boats are usually away from the coast or docked in harbors. In order to reduce false detections (notably related to waves hitting the coast), we removed the bigger boats that were in a 200m buffer zone from the coast. This decision removes potential docked ships detections in harbors.

### Clusters

Exploring the distribution of the detections is a relevant approach to spot anomalies. We plotted the data along the different information axes and visually spotted clusters. When looking at their corresponding predictions we observed some of them were false detections. In order to remove those clusters in a systematic way, we used DBSCAN with proper distance parameters for each corresponding dimension.

Using this approach we removed repetitions in time at the same location with similar geometry often corresponding to small pieces of land that were not removed using the land mask, along the following axes: latitude, longitude, length, breadth, and aspect ratio. We also removed repetitions in space with similar geometry and the same orientation which corresponds to rays of red, green or blue lights probably related to Planets' capturing instruments. These rays of light were not seen during training as we used Pléiades images.



*Figure 7. Removing Planet capturing instruments defects (aligned thin rays of lights) creating false detections, using parametrized DBSCAN for automatic cluster localization*

## Geometry

Filtering by geometry helps remove predictions that are unlikely in terms of shapes. We have drawn a limit polygon on the width by height space and removed all the predictions outside the polygon. This filtering was not applied to the "wake" class as boat wakes boxes can have a much higher diversity of shapes by nature.

## Confidence

Finally, we remove the predictions with confidence lower than the optimized confidence thresholds. Respectively 0.14 for the "boat" class and 0.29 for the "wake" class (see section 5.2 Model - Performances).

# 5 Results

## 5.1 Model evaluation

This section concerns the model evaluation on detections before data filtering, only the confidence thresholds for both classes are used.

### Test set performances

To properly evaluate the model, knowing that most of our objects are small, we have set the IoU threshold to 0.3. For large objects, a slight offset in the predicted box barely changes the IoU, while for small objects even a small shift or size mismatch drastically impacts the score. To evaluate the performances on the test set, we select the optimal confidences on the validation set that maximizes the f1-score for each class. We observe results slightly better for the class "boat" according to the f1-score.

class	confidence	precision	recall	f1-score
boat	0.14	<b>0.85</b>	0.78	<b>0.81</b>
wake	0.29	0.75	<b>0.86</b>	0.80

Table 2. Test set evaluation with IoU = 0.3 and confidence thresholds optimized on the validation set

To further investigate the "boat" class, we divided the objects into 20 quantiles (around 60 objects each) according to vessel length, ranging from the smallest (group 1) to the largest (group 20). We found that performance is lowest for the smallest vessels (group 1), then steadily improves with increasing vessel length and stabilizes around 0.8.

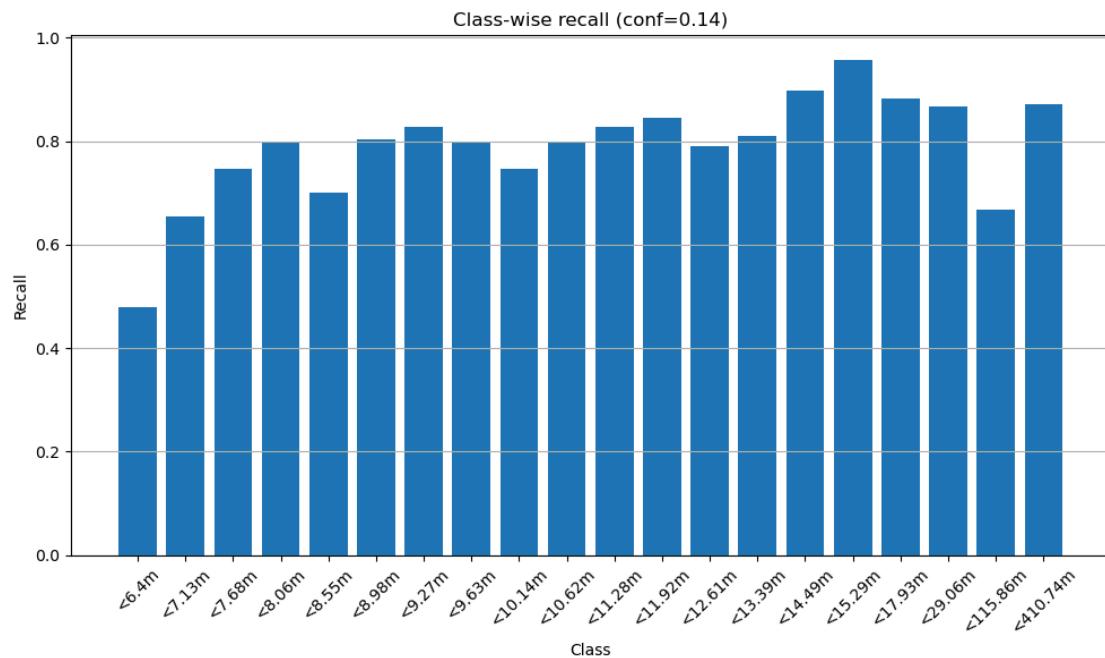


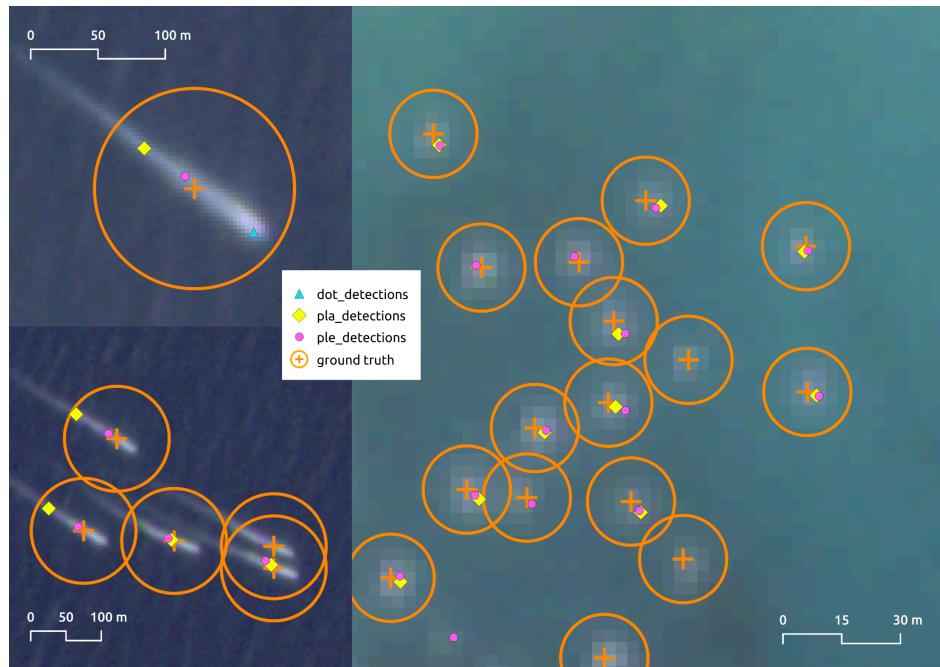
Figure 8. Model recall on the test set arranged into 20 quantiles based on boat length. The model performance tends to augment with the boat length

## Planet evaluation

In this section, we compare 3 models: PLE, Ultralytics' yolo11m-obb model pretrained on DOTAv1 that we trained on the Pléiades degraded dataset; PLA, that we trained exclusively on Planet images; DOT, Ultralytics' yolo11m-obb model that was trained on DOTAv1 dataset.

To evaluate the models we have annotated a Planet satellite image by comparing it to a Pléiades image taken within a few hours interval. As the bigger boats and moving boats are easy to spot on Planet thanks to their size or wake, Pléiades was mostly used to annotate the static smaller boats. One can consider this as a Planet image with a "real" ground truth (similar approach to (Van Etten, 2018)). We predicted boat locations with all three models using SAHI on 80 tiles of  $1024 \times 1024$  pixels for DOT and 300 tiles of  $512 \times 512$  pixels for PLA and PLE with an overlap of 0.1. We selected the optimal confidence threshold on the image for each model and matched detection center points to ground truth center points by the distance. Two distance thresholds of 100m and 15m were set respectively for ground truth objects labeled as "wake" and "boat". No other post-processing or filtering strategies were used except for the confidence threshold.

The predicted bounding boxes were reduced to their center location, and matched using Kuhn–Munkres algorithm with the above mentioned maximum distance thresholds.



*Figure 9. Detections of the DOT, PLA and PLE models, and ground truth on the test Planet image*

Model	Optimized confidence	TP	FP	FN	Accuracy	Precision	Recall	F1-score
DOT	0.0	50	140	378	0.09	0.26	0.12	0.16
PLA	0.6	303	68	125	0.61	0.82	0.71	0.76
PLE	0.1	<b>357</b>	<b>52</b>	<b>71</b>	<b>0.74</b>	<b>0.87</b>	<b>0.83</b>	<b>0.85</b>

*Table 3. Models performances on the test Planet image with individually optimized confidence threshold*

The PLE model outperforms DOT and PLA on every metric used even though it wasn't trained on Planet images. DOT is rather a generalist object detection model for satellite images, it was not designed for this specific task which explains the low performances.

We expected PLA to outperform PLE on the recall metric as the model was designed to detect white spots more frequently and because it was directly trained on Planet images. The recall-confidence and precision-confidence curves suggest that PLA is generally better for recall and PLE for precision. This difference is hidden by the choice of the optimal confidence threshold on the evaluated dataset. Indeed, doing the same evaluation with the confidence thresholds optimized on the respective validation sets of both models gives **0.84** / 0.78 recall and 0.65 / **0.90** precision for PLA / PLE, confirming our initial thoughts.

## 5.2 Database evaluation

The database evaluation consists in comparing the PLE filtered predictions (see section 4.4 Filtering) to the AIS (Automatic identification system) database, retrieved from (*Datalastic*, n.d.), and the GFW (Global Fishing Watch) database (Global Fishing Watch, 2024). The databases were evaluated over the year 2022 in the south of Corsica, area of interest number 5 (see section 4.3 Inference).

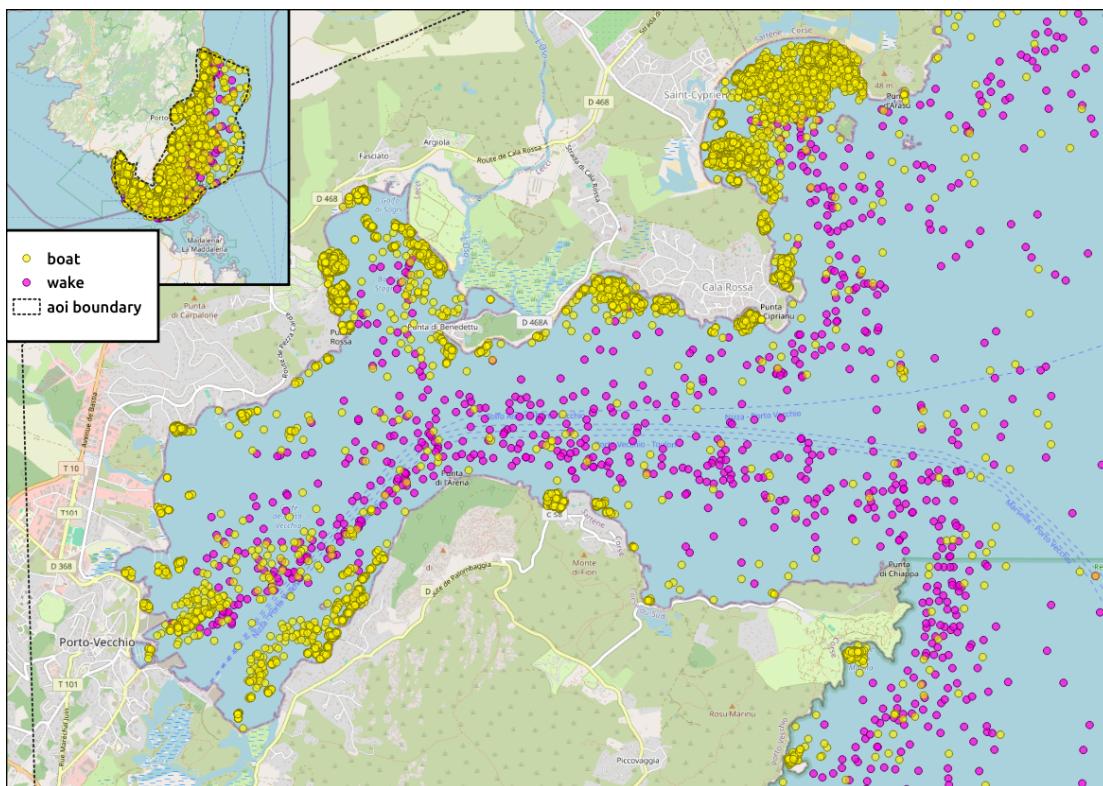


Figure 10. PLE boat detection database over 110 days in 2022 in the area of interest 5

### Rearranging the data

PLE, GFW and AIS data were rearranged so that : only the boats on common dates between 9am and 11am are kept. The time range chosen overlaps with both PlanetScope and Sentinel 1 acquisition windows. Additionally, for AIS data, if multiple boats with the same MMSI (Maritime Mobile Service Identity) at the same date remain in the database, only the first

occurrence is kept. 26 matching days were found on the study location over the year 2022, with at least one day per month and a maximum of 5 days during August.

### Detection matching

The AIS database was used as ground truth and matched to the closest GFW and PLE boats using Kuhn–Munkres algorithm with maximum distance thresholds: 100m for the “boat” class and 200m for the “wake” class. Note that AIS, GFW and PLE boats are only rarely captured at the same time resulting in poor matching in the general case. Large matching distance thresholds is an attempt to reduce this effect, nevertheless setting these values too high might generate absurd results.

The table below represents the number of matched boats (true positives) of GFW and PLE with AIS as ground truth. While PLE does not consistently achieve the highest monthly match counts, it demonstrates slightly stronger performance overall.

Month	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec	<b>Total</b>
Common days	1	3	2	2	2	3	1	5	1	2	3	1	26
AIS	5	9	7	10	59	196	122	734	88	29	19	3	1281
GFW match	<b>1</b>	<b>1</b>	0	0	2	35	14	315	10	4	2	0	384
	<b>.20</b>	<b>.11</b>	.00	.00	.03	.18	.11	.43	.11	.14	.11	.00	0.30
PLE match	0	0	0	0	<b>3</b>	<b>69</b>	<b>33</b>	<b>338</b>	<b>16</b>	4	<b>3</b>	0	<b>446</b>
	.00	.00	.00	.00	<b>.05</b>	<b>.41</b>	<b>.27</b>	<b>.46</b>	<b>.18</b>	.14	<b>.16</b>	.00	<b>0.35</b>

Table 4. GFW and PLE matched detections with AIS as ground truth. PLE has a higher matching rate overall

### Vessel count

The number of vessels detected on each common date varies similarly over the year for all 3 databases. As expected, PLE and GFW show almost systematically a higher vessel count than AIS. Also PLE estimates higher boat presence than GFW, this is probably due to the smaller boat presence much harder to detect using Sentinel-1 SAR, also they do not provide detections in a 1km buffer of the shore as the detections can be ambiguous.

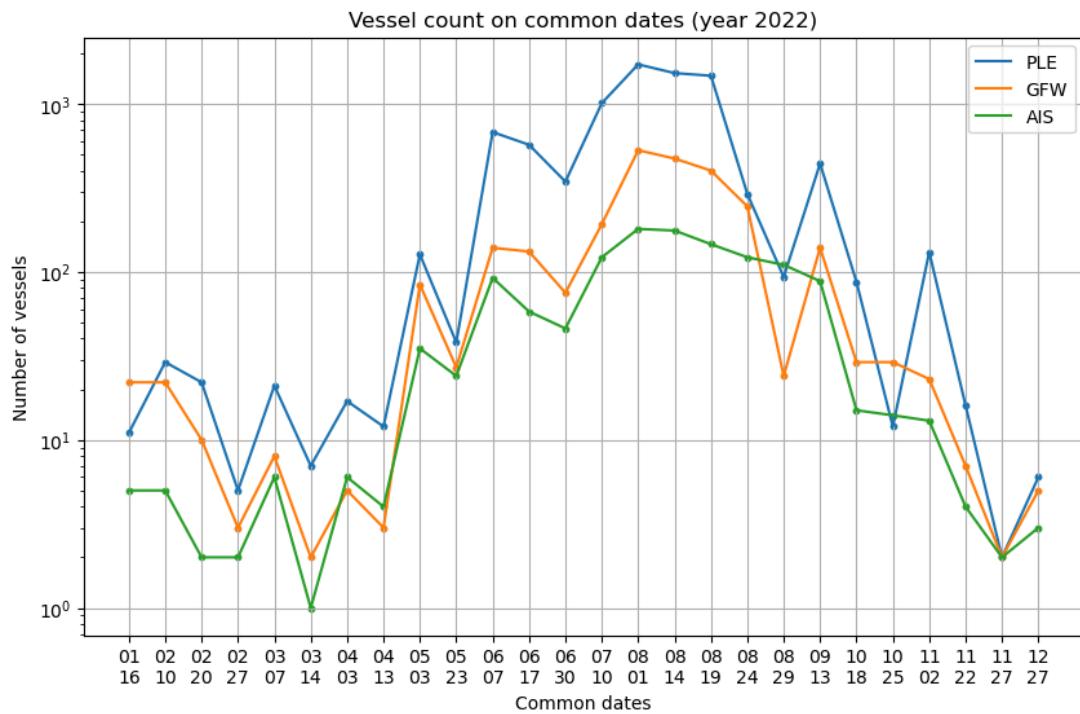


Figure 11. PLE, GFW and AIS vessels count on the common dates over the year 2022 in the study area. All three estimations vary similarly. PLE and GFW estimate a higher number of boats in most cases.

Applying a log-linear regression model to the PLE vessel count against the AIS vessel count on the 67 common dates gives an  $R^2$  of 0.90, suggesting a correspondence between the two datasets. PLE effectively captures trends in vessel activity recorded by AIS in the study area with a non-linear relationship. PLE tends to grow slightly faster than AIS (slope of 1.28), estimating generally a higher boat count.

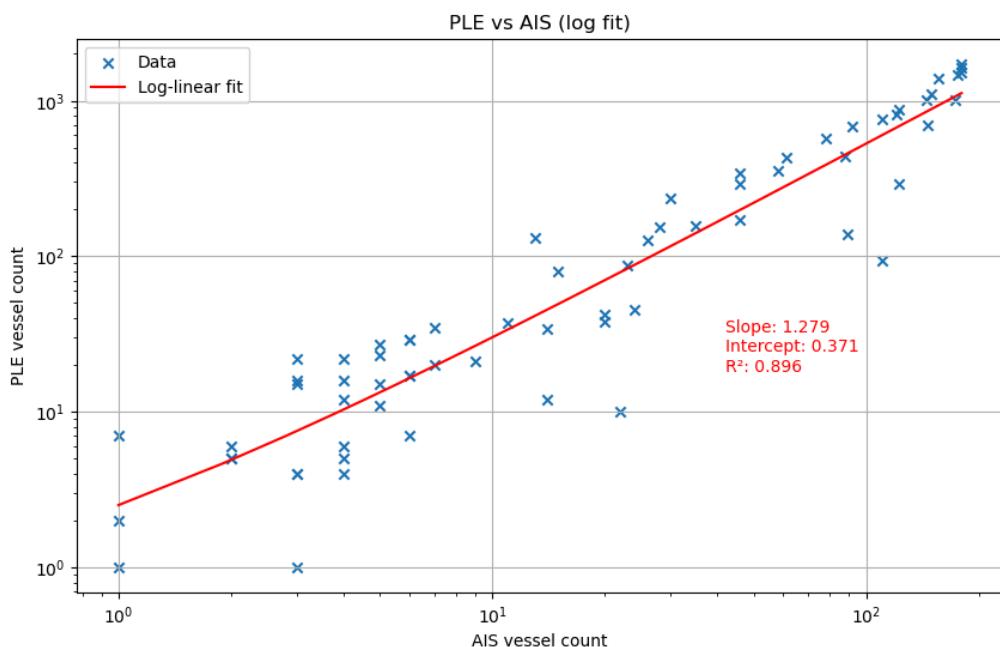


Figure 12. Log linear regression of PLE against AIS vessel counts over 67 common dates in the study area

## 6 Limitations and discussion

### Limitations of the style transfer method

The transformation applied on Pléiades images to mimic Planet visual style is a basic method that is visually convincing but can be improved. In an attempt to evaluate the transformation we have created two sets of images: one consisting of Pléiades transformed tiles, the other one with Planet tiles. We trained a YOLOv11 model to classify the images, the model achieves great classification performances on the test set after only a few epochs. We further explored the differences by encoding the images with ResNet features and reducing the dimensionality using PCA and t-SNE. Projecting the data in three dimensions revealed a clear separation between the two sets, suggesting that the images are distinguishable.

One of our colleagues, Lily Gentaz, explored the possibility of using GANs to generate the image transformation. Using GANs might sound tricky to generate images to train on as it creates new objects. In a usual case, one could train GANs by using an image from set A with its corresponding image in set B as target. This strategy doesn't work in the case of boats as corresponding images are taken within a few hours interval during which boats have moved. To address this issue, researchers proposed a model that uses boat segmentation within the GAN training loop to guide the model towards generating the boats (Zhang et al., 2024).

### Limitations of model evaluation

The model comparison against PLA and DOT suffers from the small amount of data. The Planet satellite images are manually annotated by referencing a Pléiades image captured within a short time interval. Identifying the two corresponding images is relatively difficult, the annotation process is also challenging as the boats may have moved slightly between the images. Facilitating this process to create a larger dataset would enhance the reliability of the results.

### Limitations of database evaluation

The database evaluation should be conducted on more study areas over a longer period of time. One must keep in mind that the performances are representative of this particular area over the year 2022. Additionally, an improved AIS matching algorithm should be used. Interpolating the location of the vessels based on the previous and next locations and using the orientation and speed, would improve the matching rate and enhance the reliability of the results.

### Point model for tiny boat detection

As discussed previously, the smallest boats have a recognisable normal distribution very similar to white spots often observed in microscope imagery. The Piscis deep learning algorithm designed to detect spots on microscope images was briefly tested by training a pretrained model on our dataset (Niu et al., 2024). Using Piscis is a promising approach as it showed performances a little lower than YOLOv11 on small boats with a very accurate location prediction and a higher precision. Further research must be done.

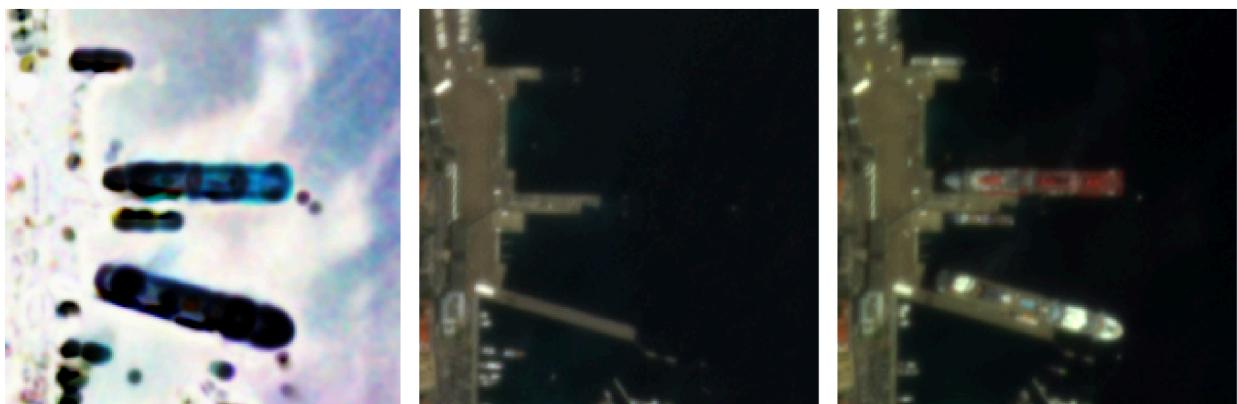
### Boat type classification using nearby context

Classifying smaller boats at a coarse image resolution can be challenging, therefore the most promising classification models use environmental raster layers as context to classify the detections and not only rely on the image layer (Paolo et al., 2024). Another approach would be to use large boxes as ground truth to consider nearby visual context, such as boat wake,

to classify smaller boats that lack visual details. One advantage is that it should be straightforward to have a naïve implementation by enlarging boxes of an existing annotated dataset.

### Boat detection by image difference

Boat detection on satellite images often rely on deep learning models which require intense labor to construct the datasets, can be biased, and are high in energy consumption during training. As an alternative solution we propose to study image differences. As boats are one of the few moving objects between images that are well georeferenced, their difference can reveal boat locations and shapes. This approach also requires further post processing to avoid false detections (e.g. landmask, cloudmask). For example an “empty” image, that has no boats and a clear view, could be constructed using multiple images of the same site (similar to a long exposure); subtracting this image to a satellite image will reveal the moving objects precisely.



*Figure 13. Structural Similarity Index Measure map (left) of two satellite images (center and right) with one day interval reveals precise boat locations.*

## 7 Conclusion

Our workflow showed to be relevant in the case of small boat detection model training. The trained model showed good performances on the test set (0.81 f1-score) with parameters optimized on the validation set. The model adapts fairly well to a real case application as it outperforms the two rival models tested on a properly annotated Planet image (with 0.85 f1-score). Additionally the database derived from the model inference has higher performances than Sentinel-1 derived global fishing watch database against AIS as ground truth. Finally the vessel count estimation shows an exponential relationship to AIS count over time, suggesting that the model variations in time are realistic, and emphasizing on the fact that AIS alone does not properly represent human pressure in the sea. Nevertheless we suggest conducting additional evaluations on other properly annotated Planet images, as well as a more fine grained AIS comparison.

## 8 References

- Ager, T. (2013). An Introduction to Synthetic Aperture Radar Imaging. *Oceanography*, 26(2).  
<https://doi.org/10.5670/oceanog.2013.28>
- AIS (Automatic Identification System) Overview*. (2021). Shipping.Nato.Int.  
<https://shipping.nato.int/nsc/operations/news/2021/ais-automatic-identification-system-overview.aspx>
- Akyon, F. C., Onur Altinuc, S., & Temizel, A. (2022). Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. *2022 IEEE International Conference on Image Processing (ICIP)*, 966–970. <https://doi.org/10.1109/ICIP46576.2022.9897990>
- Bannister, N. P., & Neyland, D. L. (2015). Maritime domain awareness with commercially accessible electro-optical sensors in space. *International Journal of Remote Sensing*, 36(1), 211–243. <https://doi.org/10.1080/01431161.2014.990647>
- Datalastic. (n.d.). Vessel Tracking API & Ship AIS Database. Retrieved August 18, 2025, from <https://datalastic.com/>
- Faudi, J., & Martin. (2018). *Airbus Ship Detection Challenge*.  
<https://kaggle.com/competitions/airbus-ship-detection>
- Gallego, A.-J., Pertusa, A., & Gil, P. (2018). Automatic Ship Classification from Optical Aerial Images with Convolutional Neural Networks. *Remote Sensing*, 10(4), 511. <https://doi.org/10.3390/rs10040511>
- Global Fishing Watch. (2024). *SAR vessel detections pipev3* [Dataset].
- Gower, J., & Skey, S. (2000). *Wind, Slick, and Fishing Boat Observations with Radarsat ScanSAR*. 21.
- Kanjir, U., Greidanus, H., & Oštir, K. (2018). Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sensing of Environment*, 207, 1–26. <https://doi.org/10.1016/j.rse.2017.12.033>
- Khanam, R., & Hussain, M. (2024). *YOLOv11: An Overview of the Key Architectural Enhancements* (No. arXiv:2410.17725). arXiv.  
<https://doi.org/10.48550/arXiv.2410.17725>
- Le Berre, I., Foulquier, E., Le Guyader, D., Iphar, C., Sahuquet, M., & Lopez, P. J. (2024). De l'emprise à l'empreinte: Cartographier la donnée AIS pour qualifier l'occupation de l'espace maritime caribéen. *Cybergeo: European Journal of Geography*. <https://doi.org/10.4000/123ic>
- Li, B., Xie, X., Wei, X., & Tang, W. (2021). Ship detection and classification from optical remote sensing images: A survey. *Chinese Journal of Aeronautics*, 34(3), 145–163. <https://doi.org/10.1016/j.cja.2020.09.022>
- Liu, S., Chen, P., & Zhang, Y. (2023). A Multiscale Feature Pyramid SAR Ship Detection Network With Robust Background Interference. *IEEE Journal of Selected Topics in*

- Applied Earth Observations and Remote Sensing*, 16, 9904–9915.  
<https://doi.org/10.1109/JSTARS.2023.3325376>
- MarineTraffic*. (n.d.). MarineTraffic: Global Ship Tracking Intelligence. Retrieved August 21, 2025, from <https://www.marinetraffic.com>
- Meares, J. (2024, October 21). *How we used AI to build our newest Open Buildings dataset*. Google. <https://blog.google/products/maps/google-open-buildings-dataset/>
- Moujahid, B.-L., Rodolphe, V., Guillaume, H., & Fablet, R. (2023). *Automated Ship Detection and Characterization in Sentinel-2 Images: A Comprehensive Approach*.
- Niu, Z., O'Farrell, A., Li, J., Reffsin, S., Jain, N., Dardani, I., Goyal, Y., & Raj, A. (2024). *Piscis: A novel loss estimator of the F1 score enables accurate spot detection in fluorescence microscopy images via deep learning*. Bioinformatics. <https://doi.org/10.1101/2024.01.31.578123>
- OpenStreetMap. (2017). *Planet dump* [Dataset]. <https://planet.osm.org/>.  
<https://www.openstreetmap.org>
- Paolo, F. S., Kroodsma, D., Raynor, J., Hochberg, T., Davis, P., Cleary, J., Marsaglia, L., Orofino, S., Thomas, C., & Halpin, P. (2024). Satellite mapping reveals extensive industrial activity at sea. *Nature*, 625(7993), 85–91. <https://doi.org/10.1038/s41586-023-06825-8>
- Pelich, R., Longepe, N., Mercier, G., Hajduch, G., & Garello, R. (2015). Performance evaluation of Sentinel-1 data in SAR ship detection. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2103–2106. <https://doi.org/10.1109/IGARSS.2015.7326217>
- PlanetScope*. (2025, July 21). *PlanetScope | Planet Documentation*. <https://docs.planet.com/data/imagery/planetscope/>
- Pléiades full archive and tasking*. (n.d.). Earth Online. Retrieved August 21, 2025, from <https://earth.esa.int/eogateway/catalog/pleiades-full-archive-and-tasking>
- Ricci, S. W., & Bohnenstiehl, D. R. (2022). Monitoring visitation at North Carolina artificial reef sites using high spatiotemporal resolution PlanetScope imagery. *Regional Studies in Marine Science*, 55, 102511. <https://doi.org/10.1016/j.rsma.2022.102511>
- Robert Hammell. (n.d.). *Ships in Satellite Imagery* [Dataset]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/61115>
- Sentinel-1 Mission*. (n.d.). Sentiwiki Copernicus. Retrieved August 21, 2025, from <https://sentiwiki.copernicus.eu/web/s1-mission>
- Sentinel-2 Mission*. (n.d.). Sentiwiki Copernicus. Retrieved August 21, 2025, from <https://sentiwiki.copernicus.eu/web/s2-mission>
- Van Etten, A. (2018). *Aircraft Detection at Planetary Scale*. <https://www.planet.com/pulse/aircraft-detection-at-planetary-scale/>

- Wen, L., Cheng, Y., Fang, Y., & Li, X. (2023). A comprehensive survey of oriented object detection in remote sensing images. *Expert Systems with Applications*, 224, 119960. <https://doi.org/10.1016/j.eswa.2023.119960>
- Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2018). DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3974–3983. <https://doi.org/10.1109/CVPR.2018.00418>
- Zhang, H., Li, H., Lin, J., Zhang, Y., Fan, J., & Liu, H. (2024). Seg-CycleGAN: SAR-to-optical image translation guided by a downstream task (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2408.05777>

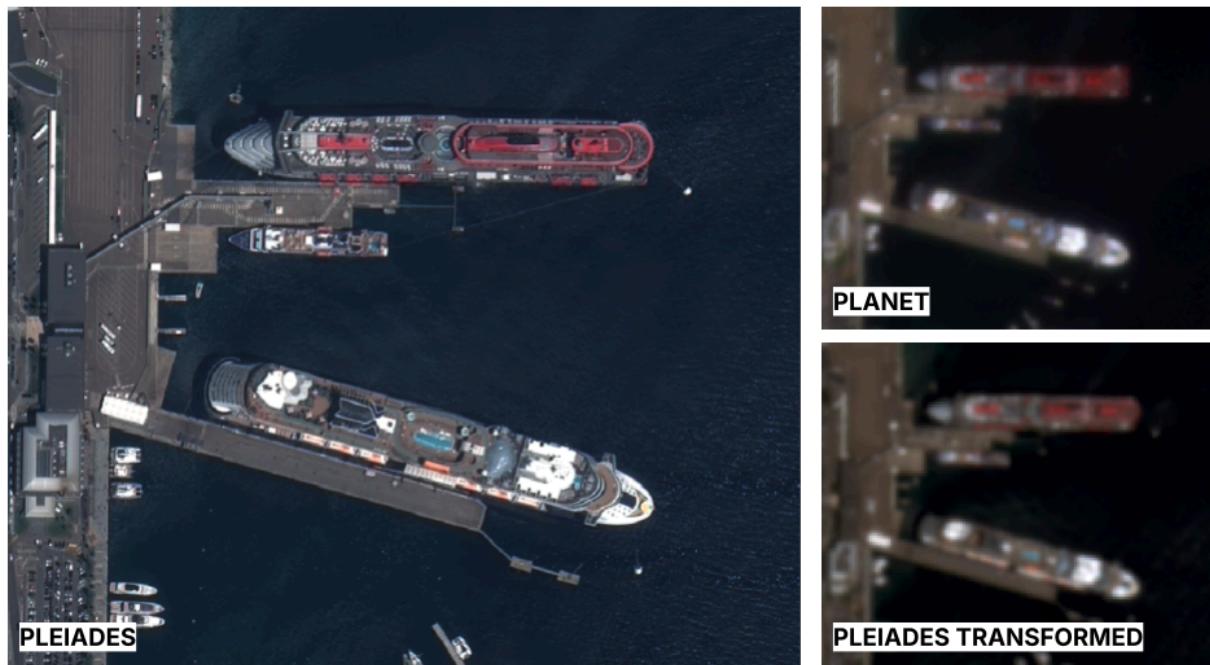
## 9 Appendix

Abbreviation	Category	Meaning
DOT	model	The yolo11m-obb model trained on DOTAv1 dataset.
PLA	model	The yolo11m-obb model trained on Planet images. At 3m/px the smallest boats are not recognisable and could be mixed with foam or pieces of land, and therefore the dataset is considered "naively" annotated.
PLE	model database	This abbreviation refers both to: <ul style="list-style-type: none"> <li>- the yolo11m-obb model trained on Pléiades transformed images</li> <li>- the boat detection database constructed using the above mentioned model with post-processing filtering of the detections</li> </ul>
GFW	database	The Global Fishing Watch boat detection database
AIS	database	The Automated Identification System boat presence database retrieved from datalastic.com

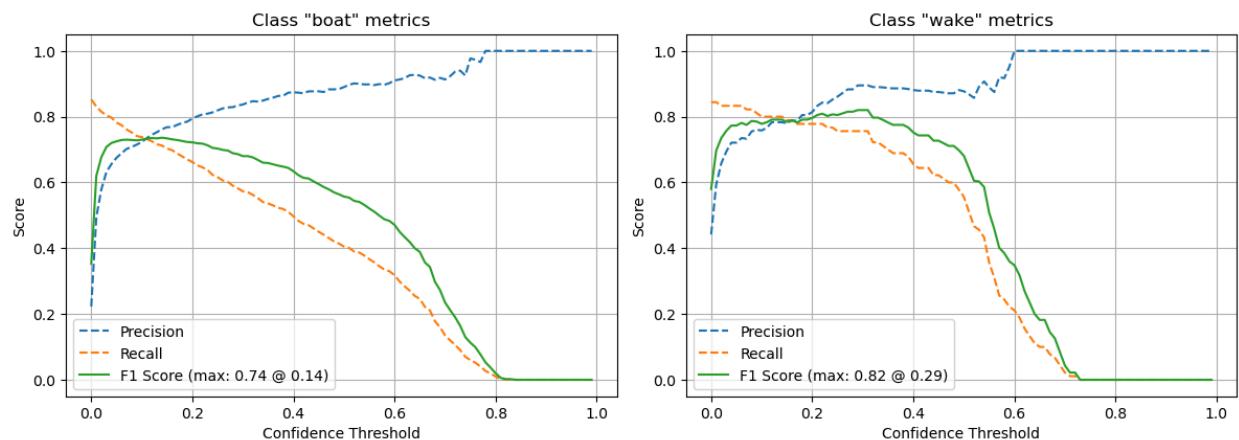
Table 5. Models and databases



Figure 14. Pléiades, Planet, and Pléiades transformed using downscaling, gaussian blur and histogram matching



*Figure 15. Pléiades, Planet, and Pléiades transformed using downscaling, gaussian blur and histogram matching*



*Figure 16. Precision, recall and f1-score against minimum confidence threshold on the validation set for PLE training with an iou of 0.3*

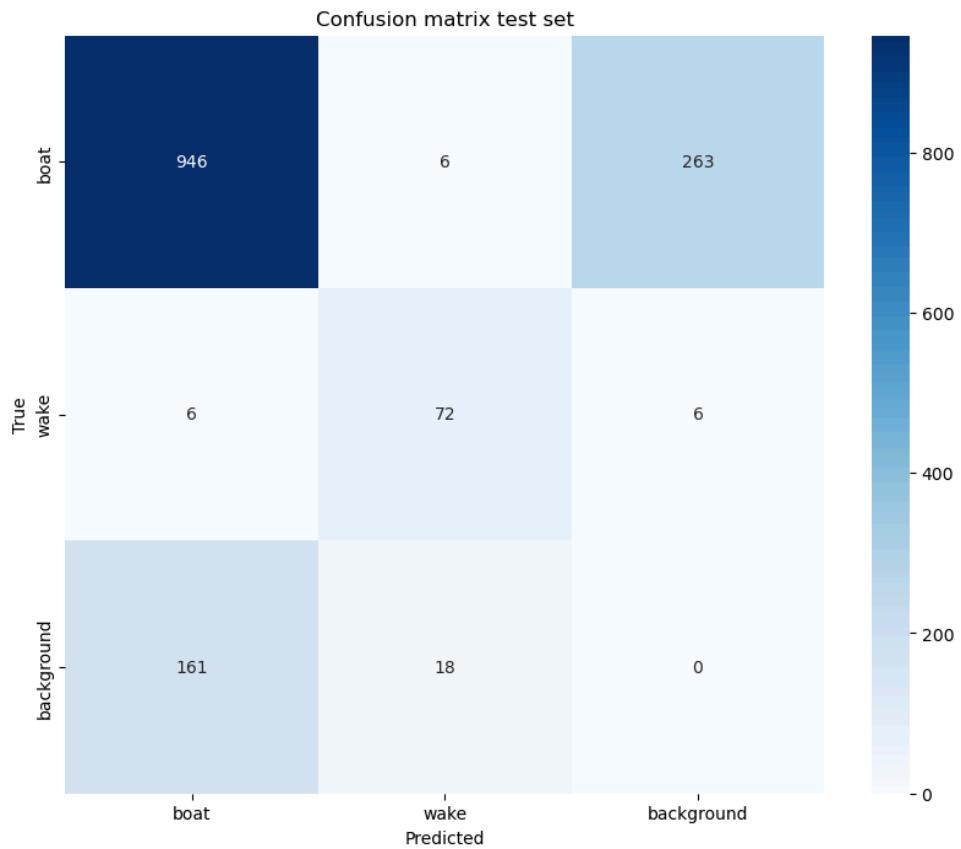


Figure 17. Confusion matrix of PLE on the test set with confidence thresholds optimized on the validation set (0.14 for the boat class and 0.29 for the wake class) and an iou of 0.3

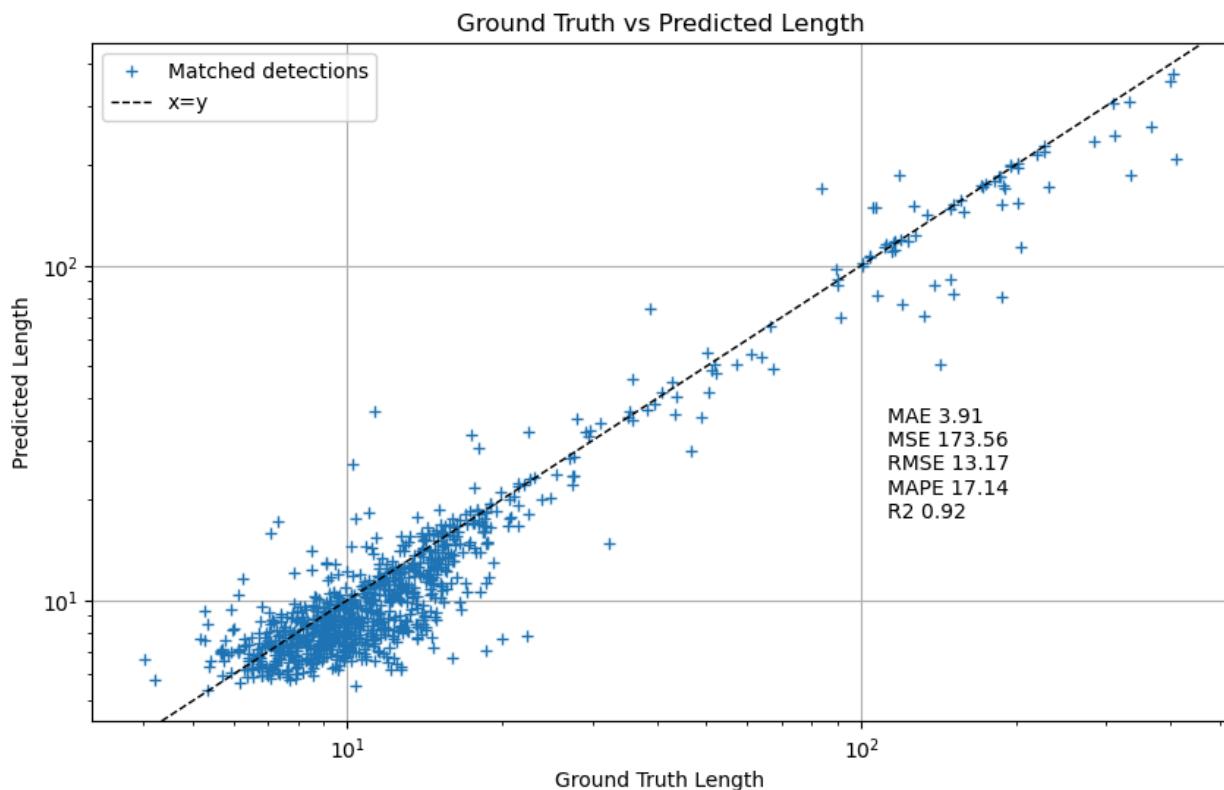


Figure 18. Ground truth length against predicted length

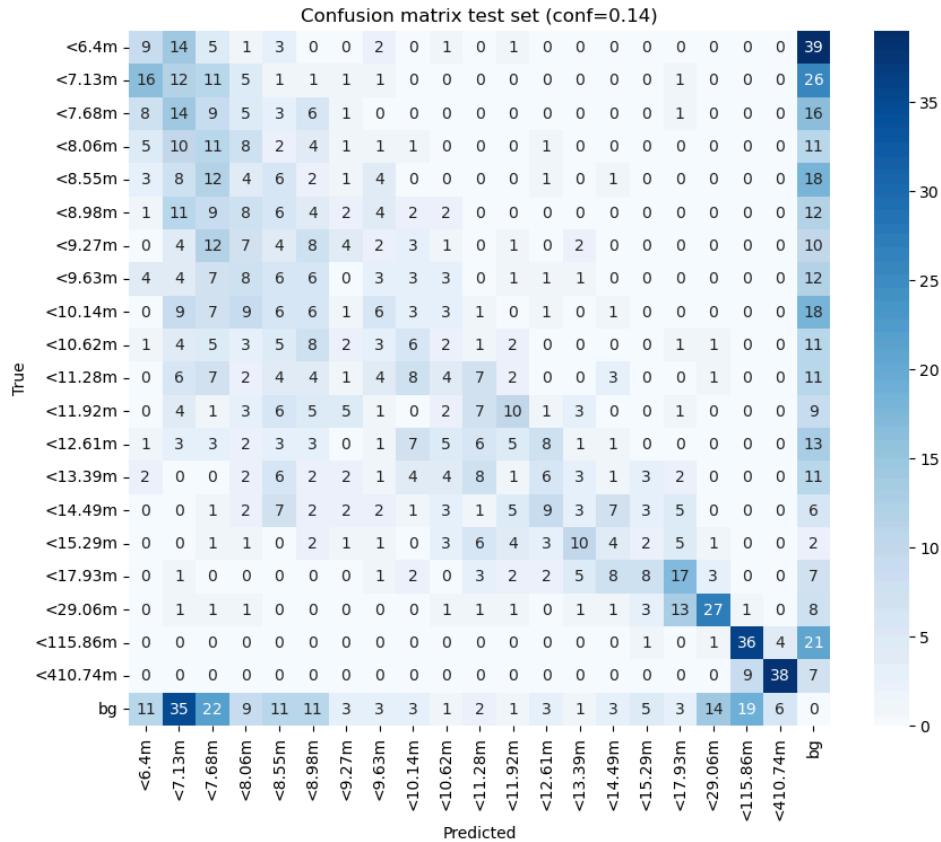


Figure 19. Confusion matrix by boat size on the test set

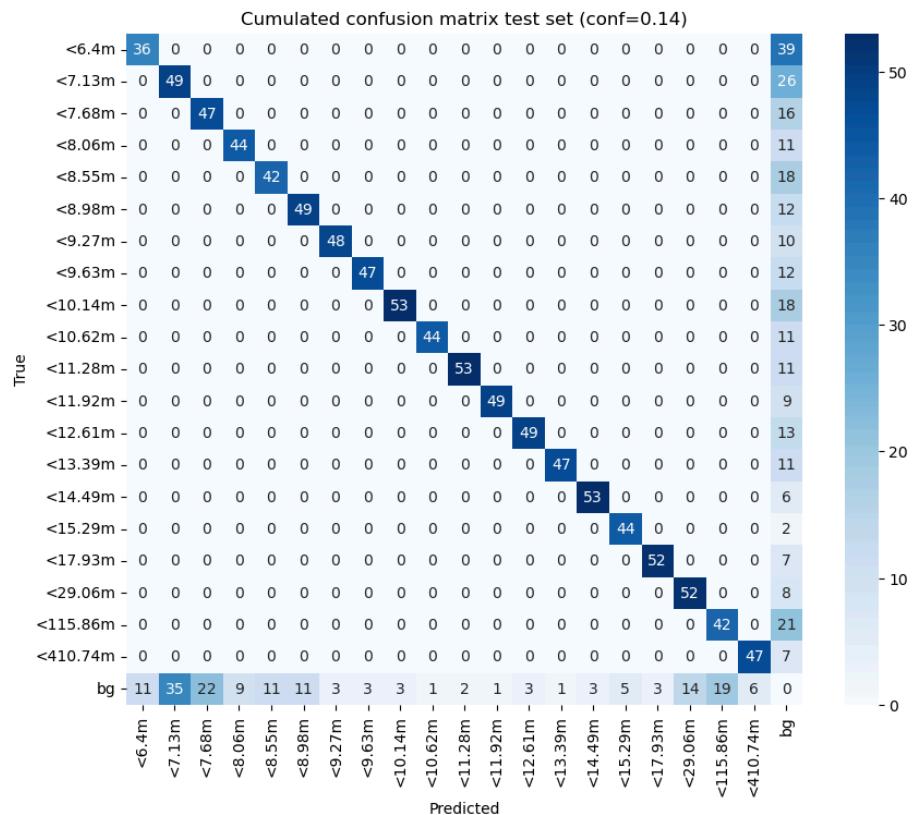
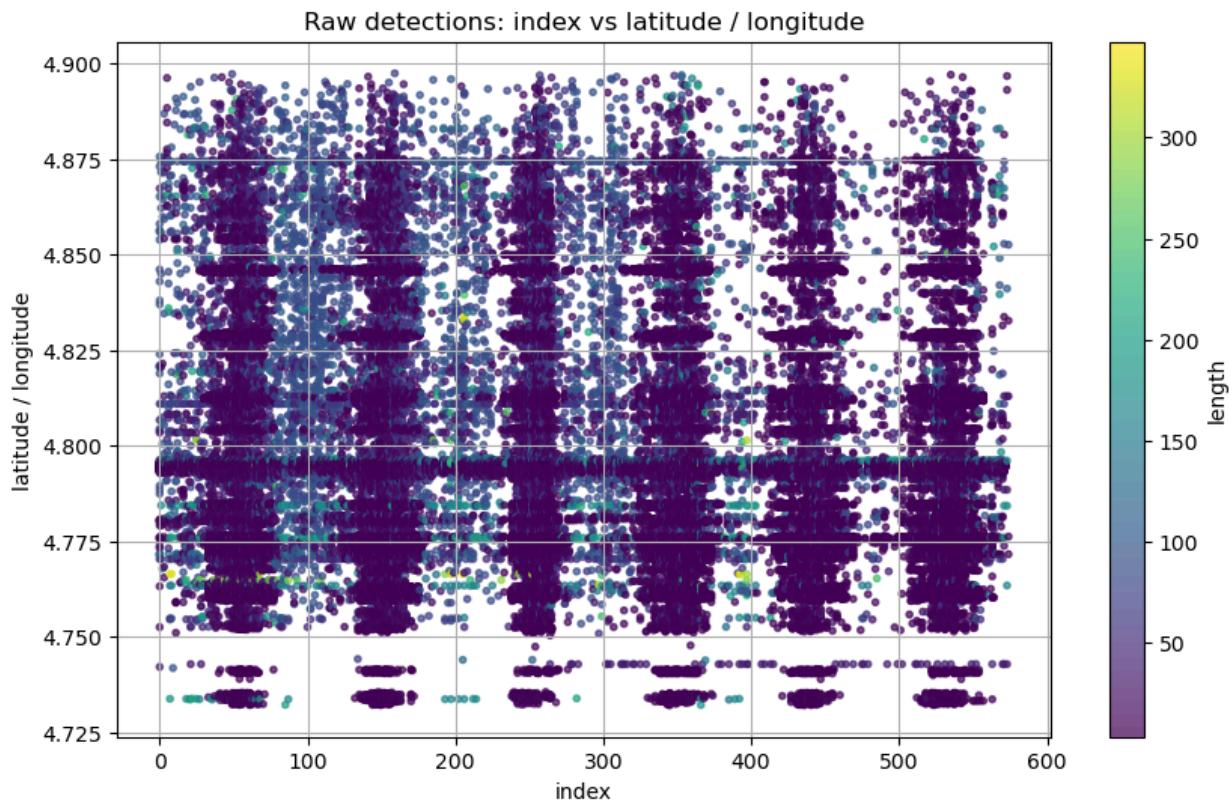
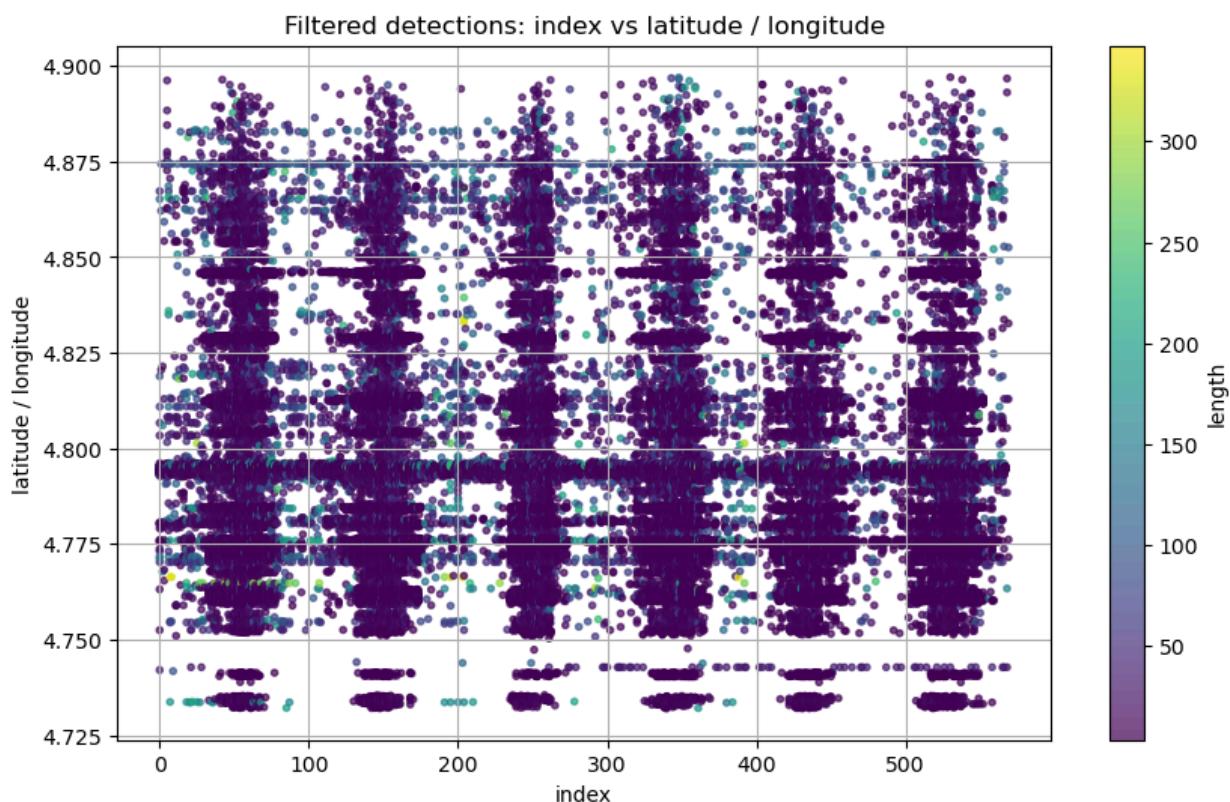


Figure 20. Cumulated confusion matrix on the test set by boat size. The detected objects are cumulated together to reveal the recall performances erasing misclassification



*Figure 21. Raw predictions on aoi 19 over 6 years with a location proxy on the vertical axis, and the date index on the horizontal axis. We observe a background noise that disappears around index 350 might be related to a satellite capturing instrument update*



*Figure 22. The background noise is removed using cluster filtering*