CHALMERS
UNIVERSITY OF TECHNOLOGY

Computer Vision Framework for Recreational Boat Speed Estimation and Corresponding Noise Mapping from Satellite Imagery

Master's thesis in Mechanics and Maritime Sciences

QI WANG, RUOYI TAO

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Computer Vision Framework for Recreational Boat Speed Estimation and Corresponding Noise Mapping from Satellite Imagery

QI WANG, RUOYI TAO

Department of Mechanics and Maritime Sciences
Division of Marine Technology
Chalmers University of Technology
Gothenburg, Sweden 2026

Computer Vision Framework for Recreational Boat Speed Estimation and Corresponding Noise Mapping from Satellite Imagery
QI WANG & RUOYI TAO

Supervisor: Wengang Mao, Department of Mechanical Engineering
Industrial Supervisor: Dag Glebe, IVL Swedish Environmental Research Institute
Examiner: Wengang Mao, Department of Mechanical Engineering

Master's Thesis 2026
Department of Mechanics and Maritime Sciences
Division of Marine Technology
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Abstract

Recreational boats are an unnegligible source of underwater radiated noise in coastal areas, especially during the summer season. Traditional noise mapping methods mainly rely on AIS-based ship position and speed information, which works better for merchant ships than for small recreational vessels. However, most small recreational boats are not equipped with reliable Automatic Identification System. This creates a data gap for vessel monitoring and underwater noise assessment. This thesis investigates whether AI-based computer vision methods and machine learning can be used to detect recreational boat wakes from optical satellite imagery and support speed-based noise mapping.

The proposed framework is based on Sentinel-2 L2A optical imagery with 10 m spatial resolution. A wake detection dataset was built from satellite images and annotated for YOLO-based object detection. The detection model was designed for small and weak wake features, using multi-scale inference and sliced image prediction. To support speed estimation, detected wakes were linked for ships with AIS records using geolocation and imaging timestamps. The matched samples were then processed into a speed-wake dataset, including denoised, feature-enhanced and augmented wake images. A CNN-based regression model was used to estimate vessel speed from visual wake features.

The results show that YOLO-based models, as machine learning based computer vision methods, can be applied to detect wake-like vessel traces that are missing from AIS-based observations. In one case study, the YOLO-based detector identified substantially more vessel wake objects than the available AIS records in the same satellite scene. A high percentage of the vessel wake objects identifiable with the human eyes were detected. The detected objects were initially matched with AIS records to extract vessel speed information, where only a small proportion of the corresponding AIS records were found. The speed of the vessel wake objects without available AIS records were then estimated using a CNN-based speed regression model. The complete vessel speeds were further used to produce a simplified underwater noise map, using an empirical noise model and an acoustic propagation model. These results suggest that AI-based wake detection and vessel speed estimation can provide a useful complementary data source for monitoring recreational boat activity and assessing its potential noise impact in coastal waters. The approach is still limited by image resolution, cloud cover, sea clutter, wake overlap, and the availability of reliable speed labels. However, it provides a practical basis for future research on small-vessel monitoring from remote sensing data.

Keywords: recreational boats, wake detection, vessel speed prediction, computer vision, machine learning, YOLO, regression network, underwater noise map

Acknowledgements

We would like to express our sincere gratitude to our supervisor, Wengang Mao, for his continuous guidance throughout this thesis. His timely feedback and direction helped us stay on track and complete each phase of the project successfully.

We are also grateful to our industrial supervisor, Dag Glebe, for providing valuable insights into the practical applications of our work.

We would like to thank our families and friends for their unwavering support and encouragement throughout this journey.

Finally, we appreciate the collaborative experience and mutual support during this project.

Qi Wang, Ruoyi Tao, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AI	Artificial Intelligence
AIS	Automatic Identification System
CNN	Convolutional Neural Network
CPU	Central Processing Unit
GPU	Graphics Processing Unit
HBB	Horizontal bounding box
IoU	Intersection-over-Union
MAE	Mean Absolute Error
ML	Machine Learning
MMSI	Maritime Mobile Service Identity
MSE	Mean Squared Error
NMS	Non-Maximum Suppression
NN	Neural Network
ProgLoss	Progressive Loss Balancing
PSA	Partial Self-Attentio
RAM	Random Access Memory
SAHI	Slicing Aided Hyper Inference
SAR	Synthetic Aperture Radar
SOG	Speed Over Ground
SPL	Sound Pressure Level
STAL	Small-Target-Aware Label Assignment
URN	Underwater Radiated Noise
YOLO	You Only Look Once

Contents

List of Figures

List of Tables

1

Introduction

Underwater noise from vessels is an important environmental concern in coastal waters. Existing noise mapping approaches mainly rely on Automatic Identification System (AIS) data to describe vessel traffic and estimate noise emissions [1]. This works well for larger commercial ships, which usually transmit reliable AIS messages. However, many recreational boats are not equipped with AIS, or their AIS records are incomplete [2].

This thesis investigates how can optical satellite imagery and modern ML methods be used to reduce this data gap. The study focuses on detecting wakes of recreational boats from Sentinel-2 true colour imagery with computer vision methods, linking detected wakes with AIS records when available, estimating vessel speed from wake features, and demonstrating a simplified noise mapping case study. The aim is to explore whether machine learning based wake object detection and vessel speed estimation can provide complementary information for small-vessel monitoring.

1.1 Background

Underwater radiated noise (URN) from vessels is a major component of anthropogenic noise in marine environments. It is mainly generated by propeller cavitation, machinery, and hydrodynamic flow around the hull. Continuous vessel noise can affect marine mammals, fish, and other aquatic organisms by masking biological signals, disturbing behaviour, and changing habitat use [3]. For this reason, underwater noise has become an important topic in marine environmental monitoring and policy.

Current ship noise mapping methods are mainly developed for commercial shipping. Large-scale projects, such as the EMSA NAVISON project [1], estimate shipping sound maps by combining AIS vessel tracks, source level models, environmental data, and sound propagation models. In NAVISON, sound maps are calculated for European seas and are commonly analysed in the 63 Hz and 125 Hz frequency bands, which are relevant for low-frequency continuous underwater noise assessment. The workflow shows that AIS-based vessel traffic data can be an effective basis for large-scale noise mapping when vessel tracks and ship information are available.

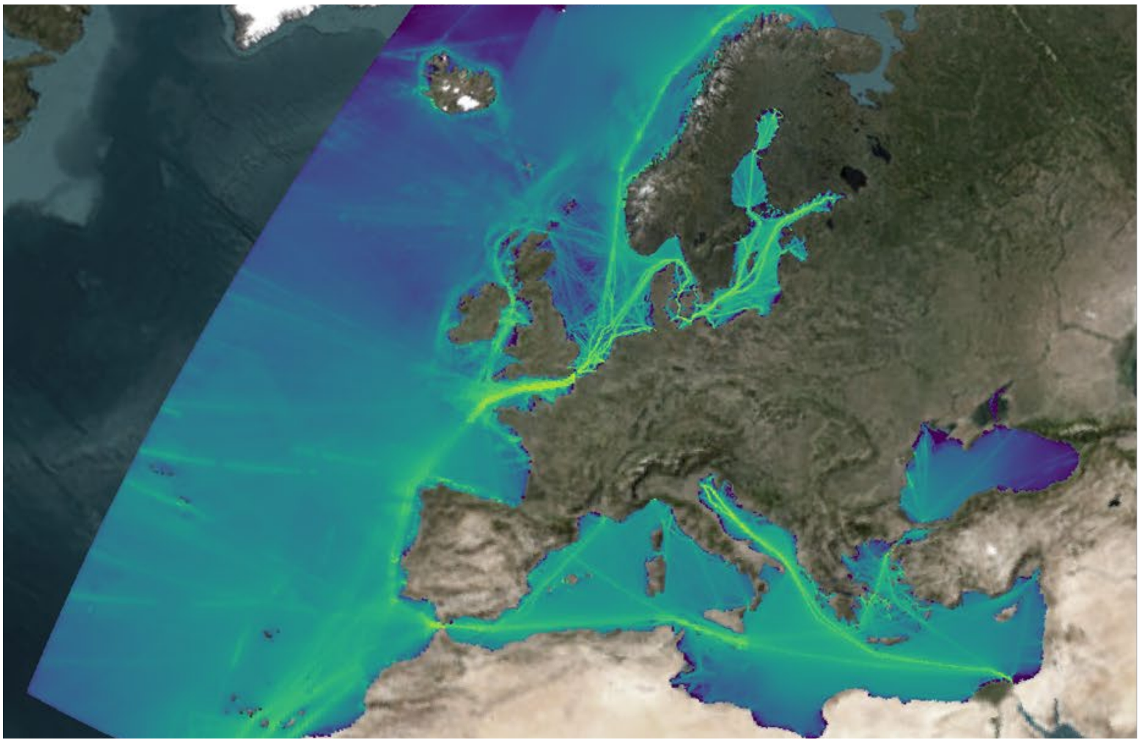


Figure 1.1: An AIS-based underwater noise map for European waters. Source: EMSA, NAVISON Final Report [1].

However, this AIS-based approach has a clear limitation for small recreational boats. Many recreational boats are not equipped with AIS transponders, and some transmit incomplete or low-quality AIS messages. This means that their positions, speeds, and trajectories are often missing from conventional vessel traffic datasets. The problem is especially relevant in coastal regions, where recreational boats are common during the summer season and often operate close to shore, islands, harbours, and sensitive marine habitats [4]. These boats can travel at relatively high speed and generate visible wakes, but their contribution is difficult to represent in traditional noise maps.

Remote sensing offers a possible way to observe vessel activity beyond AIS coverage. Optical satellite imagery can cover large coastal areas and may capture wake patterns produced by moving vessels. Sentinel-2 provides multispectral optical imagery with 10 m spatial resolution in the visible bands, which can be used to produce true colour images of coastal waters. Although this resolution is often insufficient for reliable direct detection of small recreational boats, their wakes can appear as longer linear or curved structures on the sea surface [5]. Therefore, wake detection can be used as an indirect method for identifying small-vessel activity from satellite images [6].

Recent advances in computer vision make this approach more practical. Object detection models such as YOLO can be trained to detect visual patterns in complex scenes, including small or weak targets in remote sensing images. In this thesis, YOLO-based wake detection is used to identify possible recreational boat wakes from

Sentinel-2 true colour imagery. When AIS records are available, detected wakes can be linked with vessel records using geolocation and image acquisition time. This makes it possible to build a speed-wake dataset and explore whether visual wake features can be used to train a CNN-based speed regression model to address the incompleteness of AIS records [7]. The complemented vessel speed information can then be used together as input for a simplified underwater noise mapping case study [8].

1.2 Related Work

Existing underwater noise mapping studies commonly rely on vessel traffic data, ship source models, environmental inputs, and acoustic propagation models. Large-scale projects such as NAVISON and JOMOPANS have demonstrated that AIS-based ship tracks can be used to estimate shipping noise over large sea areas [1, 9]. In these frameworks, hydrophone measurements are mainly used for validation and uncertainty assessment, while the spatial sound maps are produced through modelling. This approach is effective for large commercial vessels, but its reliability depends on the completeness of vessel traffic data. Therefore, it is less suitable for recreational boats and other small vessels that are often missing from AIS records.

Satellite-based wake detection provides a possible complementary observation method for such vessels. Traditional ship wake detection studies have used physically interpretable methods such as Radon transform, Hough transform, sparse regularization, and spectral analysis, especially for SAR imagery [10, 5]. These methods can extract wake geometry and support speed estimation, but their performance is often affected by sea state, image noise, wake visibility, and parameter selection. More recent studies have applied deep learning to ship and wake detection. CNN-based and YOLO-based methods have been used for SAR and optical satellite images, including TanDEM-X and Sentinel-2 data [11, 12, 13, 14]. These studies indicate that learning-based detectors are suitable for maritime wake detection, particularly when wakes are more visible than the vessel hull.

Several studies have further investigated wake-based speed estimation. Classical approaches relate Kelvin wake geometry and wavelength to vessel speed, while recent methods combine object detection, AIS-labelled data, and frequency-domain wake analysis [15, 16]. These studies show that satellite-observed wakes can provide useful information about vessel motion. However, their applicability is still limited by wake visibility, image quality, wake curvature, and the availability of suitable labelled data.

Research area	Representative studies	Relevance to this thesis
Underwater noise mapping	NAVISON [1]; JOMOPANS [9]	AIS-based ship tracks are widely used for large-scale underwater noise mapping, but this approach depends on the completeness of vessel traffic data.
Satellite-based wake detection	Mazzeo et al. [10]; Zilman et al. [5]; Kang and Kim [11]; Liu and Deng [12]	Satellite imagery can be used to detect ship wakes and extract motion-related information, but wake visibility, sea state, image noise, and parameter selection can affect performance.
YOLO-based maritime detection	Magalhães et al. [13]; Qiu et al. [14]	YOLO-based models have shown strong potential for vessel and wake detection in satellite imagery, including Sentinel-2 and wake-specific datasets.
Wake-based speed estimation	Gao et al. [15]; Haakman et al. [16]	Wake features can be used to estimate vessel speed, supporting the idea of linking satellite-observed wakes with speed-related maritime analysis.

Table 1.1: Summary of related studies

Table 1.1 summarizes the main studies related to underwater noise mapping, satellite-based wake detection, and wake-based speed estimation.

Overall, previous work provides three foundations for this thesis. First, AIS-based ship tracks are widely used in underwater noise mapping, but AIS records are incomplete for recreational boats [17]. Second, satellite imagery can provide complementary observations of vessel wakes when small vessels are difficult to detect directly. Third, ML based approaches, namely YOLO-based detection and wake-based speed estimation, have shown promising results. However, limited work has connected Sentinel-2 true colour wake detection of recreational boats with AIS-based speed labelling and simplified underwater noise mapping. This thesis addresses this gap by combining Sentinel-2 wake detection, AIS-based speed labelling, speed estimation from wake features, and speed-based noise mapping for recreational boat activity in Swedish coastal areas.

1.3 Objectives

The main research problem of this thesis is whether wake information from satellite images can help reduce the data gap caused by incomplete AIS coverage for small vessels with the use of ML-based methods. Based on this objective, the thesis addresses the following research questions:

- Can recreational boat wakes be detected from Sentinel-2 true colour imagery using a YOLO-based object detection model?
- How well does the wake detection model perform under different coastal imaging conditions, such as varying cloud cover, sea clutter, illumination, and wake visibility?
- Can detected wakes be spatially and temporally matched with AIS records to construct a speed-wake dataset?
- Can visual wake features be used to estimate vessel speed from satellite image crops with a CNN-based speed regression model?
- How can the estimated vessel speeds be used to produce a simplified underwater noise map for a coastal case study?

1.4 Scope and Limitations

This thesis focuses on the development and evaluation of a machine learning based workflow for recreational boats including wake detection, machine learning based vessel speed estimation, and noise mapping. The study is restricted to visible wake signatures observed in coastal environments from Sentinel-2 optical imagery and does not consider radar imagery or other remote sensing modalities.

This thesis has several limitations related to the image data, annotation process, satellite acquisition process, and AIS data quality:

- Since this study uses image-based object detection, the wake detection results depend on the quality and spatial resolution of the satellite images. Weak, narrow, or low-contrast wakes may be difficult to detect in Sentinel-2 true colour imagery.
- Cloud cover can reduce the quality of the satellite images. Clouds, cloud shadows, haze, and bright reflections may obscure wakes or create wake-like patterns, which can lead to missed detections or false detections.
- The wake detection dataset was created and annotated manually. Therefore, the dataset may contain human annotation errors, such as inaccurate bounding boxes, missed wakes, or inconsistent labelling decisions.
- Sentinel-2 imagery is not acquired instantaneously over the whole scene. The exact time when a wake is formed may differ from the nominal satellite ac-

quisition time by several seconds or more. This time difference can introduce uncertainty when matching detected wakes with AIS records.

- AIS data may contain missing, delayed, or inaccurate records. This is especially relevant for recreational boats, since many of them are not equipped with AIS or only provide incomplete information. As a result, some detected wakes cannot be matched with reliable vessel speed labels.

- The CNN regression model relies on paired wake–speed samples obtained through matching YOLO predictions with AIS records. However, vessels travelling at extremely low or high speeds were relatively rare in the available AIS observations. This causes the imbalance of AIS-derived training samples across different vessel speed ranges.

- The underwater noise map in this thesis is a simplified case study. It mainly uses estimated vessel speed as input and does not include a full acoustic propagation model, detailed vessel source modelling, bathymetry, sediment type, or sound speed profile.

Consequently, the results should be interpreted primarily as a feasibility study demonstrating the potential of applying a machine learning framework for the study of recreational boats activity.

1.5 Contributions

This thesis contributes to recreational boat monitoring by combining Sentinel-2 optical satellite imagery, computer vision, AIS matching method, CNN-based speed regression and underwater noise mapping. First, a Sentinel-2 wake detection dataset was created from true colour satellite images of Swedish coastal areas. The dataset includes images collected under different seasonal, weather, and cloud cover conditions, and provides manually annotated wake samples for training and evaluation. In addition, a speed-wake dataset was constructed by matching detected wake samples with AIS records using image geolocation and acquisition time. This dataset provides speed labels for wake image crops and supports the development of image-based speed estimation methods.

Based on this diverse dataset, we trained and evaluated the YOLO-based wake detection models. Different model settings and detection strategies were compared to analyse their performance on small and weak wake features in coastal satellite imagery. A CNN-based regression model was then developed to estimate vessel speed from visual wake features. Finally, the estimated vessel speeds were used to produce a simplified underwater noise mapping case study. These contributions demonstrate a complete end-to-end machine learning framework from satellite-based wake detection to speed estimation and noise map generation, providing a complementary method for small-vessel monitoring when AIS coverage is incomplete.

2

Theory

This chapter provides the theoretical foundation for the proposed framework. It first introduces the basic concepts of computer vision and object detection. Image preprocessing methods for noise reduction and feature enhancement are then reviewed. The relationship between wake morphology and vessel speed is discussed, followed by the theoretical basis of image-based regression using convolutional neural networks. Finally, the chapter introduces the basic concepts of underwater radiated noise and noise mapping.

2.1 Computer Vision Theory

Computer vision aims to extract meaningful information from images and videos. Unlike numerical methods that mainly process structured data, such as measurements or time series, computer vision operates directly on visual data represented by pixels. It enables computers to identify patterns, objects, and regions of interest in an image.

Object detection is a major task in computer vision. It determines both the location and the category of target objects, usually by outputting bounding boxes and confidence scores. In this thesis, ship wake detection is formulated as an object detection problem, where wake regions are identified and localized from satellite imagery.

Method	Main principle	Characteristics
Traditional methods	Use hand-crafted features, such as templates, edges, or gradients, to detect target regions	Simple and interpretable, but sensitive to complex backgrounds
Faster R-CNN	Generates candidate regions and then classifies and refines them	High accuracy, but slower inference
SSD / RetinaNet	Directly predict object classes and bounding boxes from feature maps	Faster than two-stage methods
YOLO	Predicts object locations and classes in a single forward pass	Fast and suitable for large-scale image processing

Table 2.1: Comparison of representative object detection methods

Object detection methods have developed from traditional approaches based on manually designed features to modern deep-learning-based models. Traditional methods are usually more interpretable, but their performance can be limited in complex scenes. Deep learning methods can automatically learn visual features from training data and have become the dominant approach in modern object detection. A comparison of representative methods is shown in Table 2.1. The following sections introduce the related theory of deep-learning-based computer vision, using YOLO as the main example.

YOLO (You Only Look Once) is a family of single-stage object detection models. It predicts object locations and classes directly from the input image in one forward pass, making it efficient for large-scale image analysis [18]. In this thesis, YOLO is used to detect visible wake patterns of recreational boats in Sentinel-2 true-colour imagery.

As shown in Figure 2.1, Modern YOLO models usually follow a backbone–neck–head architecture. The backbone extracts hierarchical image features, the neck fuses multi-scale features, and the detection head outputs bounding boxes, class probabilities, and confidence scores.

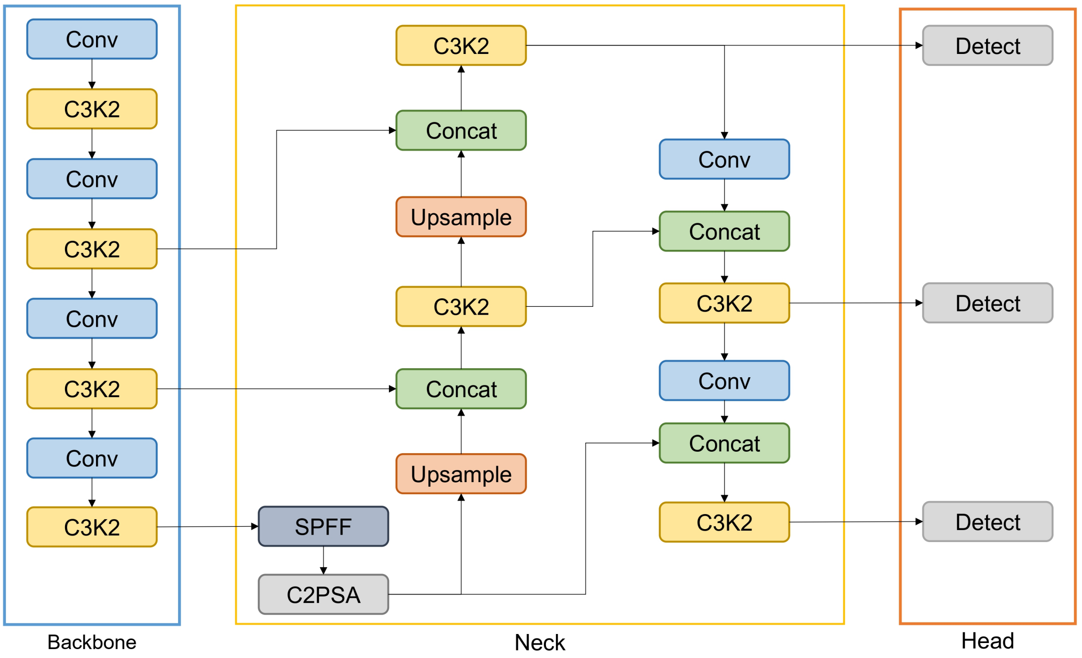


Figure 2.1: YOLO architecture

The YOLO series is a family of single-stage object detection models designed for fast and efficient detection. From YOLOv1 to later versions such as YOLOv5, YOLOv8, and YOLO11, the framework has gradually improved multi-scale feature extraction, training strategies, model scalability, and deployment efficiency. In this thesis, YOLOv26s is selected as the main wake detection model. YOLOv26 introduces several improvements, including end-to-end NMS-free inference, Progressive Loss, Small-Target-Aware Label Assignment, and the MuSGD optimizer. These features are suitable for Sentinel-2 wake detection because the target wakes are usually small, elongated, and located in complex coastal backgrounds.

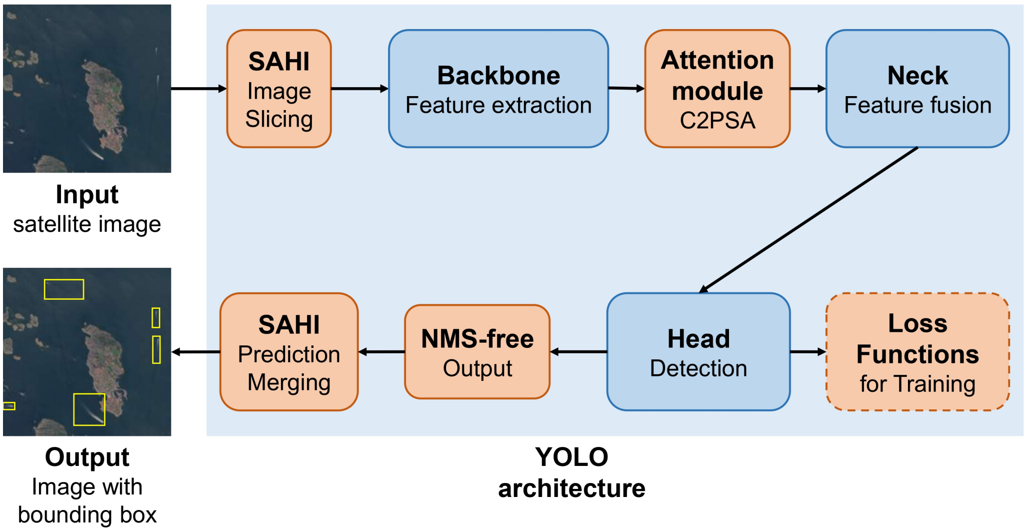


Figure 2.2: Overall workflow of the YOLO-based detection method

Figure 2.2 presents the overall workflow of the YOLOv26-based wake detection process, showing the connection between image slicing, feature extraction, attention-based feature enhancement, feature fusion, detection output, loss-based training, and prediction merging. The following subsections further introduce these modules and their roles in the proposed detection framework.

2.1.1 Attention Mechanism

Attention mechanisms are used in YOLO models to enhance feature representations by emphasizing informative spatial regions and feature channels. In YOLOv26, PSA-related modules, such as PSABlock and C2PSA, are introduced to improve global context modelling with relatively low computational cost [19]. This is useful for wake detection, since wake features in Sentinel-2 images are often weak, narrow, and easily confused with sea-surface textures.

As shown in Figure 2.3, C2PSA enhances the feature map before it is passed to the neck for multi-scale feature fusion. This helps the detector strengthen wake-related features before the final bounding box prediction.

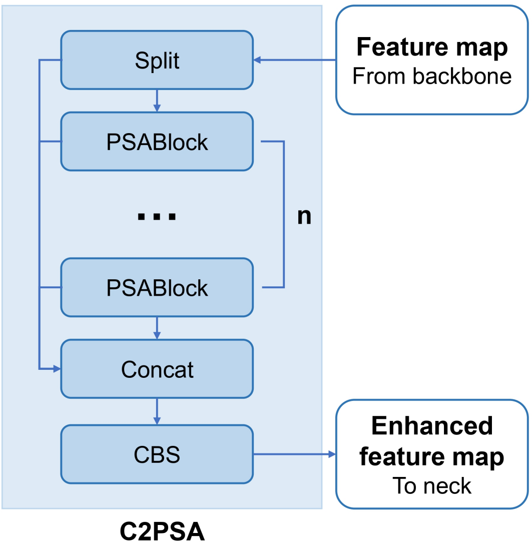


Figure 2.3: C2PSA architecture

In the attention module, the input feature map is projected into query (Q), key (K), and value (V) representations. The attention output is computed as

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2.1}$$

where Q, K, and V denote the query, key, and value matrices, and d_k is the dimension of the key vectors.

2.1.2 Loss Functions

The training objective of a YOLO detector is usually defined as a combination of localisation, classification, and objectness losses:

$$\mathcal{L}_{total} = \lambda_{box}\mathcal{L}_{box} + \lambda_{cls}\mathcal{L}_{cls} + \lambda_{obj}\mathcal{L}_{obj}, \tag{2.2}$$

where $\mathcal{L}_{box}$, $\mathcal{L}_{cls}$, and $\mathcal{L}_{obj}$ represent bounding box localisation, class prediction, and objectness losses, respectively. The coefficients λ_{box}, λ_{cls}, and λ_{obj} control their relative weights.

YOLOv26 further introduces Progressive Loss Balancing (ProgLoss) and Small-Target-Aware Label Assignment (STAL) to improve training stability and small-object detection [18, 20]. This is relevant for this thesis because boat wakes in Sentinel-2 imagery are often narrow, weak, and small relative to the full image.

2.1.3 NMS and NMS-Free Detection

Non-maximum suppression (NMS) is a common post-processing step in object detection. It is used to reduce duplicate detections when several overlapping bounding boxes are predicted for the same object. This is especially relevant for wake detection, since vessel wakes are often long, narrow, and visually ambiguous, which can lead to multiple overlapping predictions for a single wake.

For two predicted bounding boxes B_i and B_j, their overlap is measured by intersection over union (IoU):

$$IoU(B_i, B_j) = \frac{|B_i \cap B_j|}{|B_i \cup B_j|}. \tag{2.3}$$

During NMS, predictions are sorted by confidence score, and boxes with high overlap with a selected high-confidence box are suppressed:

$$B_j = \begin{cases} \text{suppressed}, & IoU(B_i, B_j) > \tau_{\text{NMS}}, \\ \text{kept}, & IoU(B_i, B_j) \leq \tau_{\text{NMS}}, \end{cases} \tag{2.4}$$

where τ_{NMS} is the IoU threshold. In dense coastal scenes, this threshold can affect the final wake detection result. A lower threshold removes duplicate boxes more aggressively, while a higher threshold may preserve nearby wakes but also leave repeated detections.

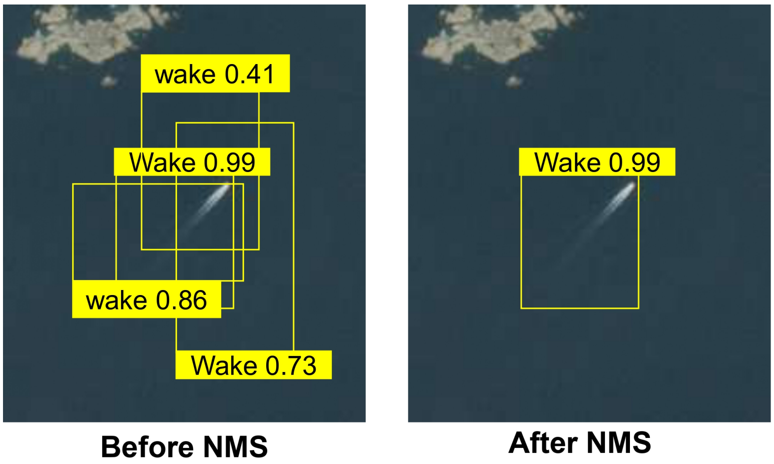
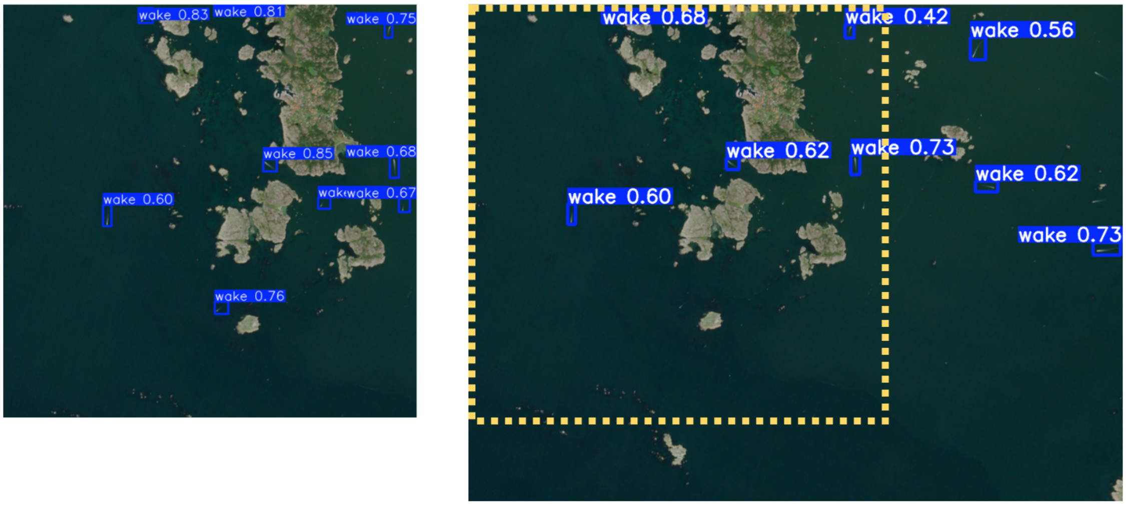


Figure 2.4: Illustration of non-maximum suppression in wake detection.

YOLOv26 also supports NMS-free detection through its one-to-one prediction head. According to the Ultralytics documentation, this head produces end-to-end predictions without requiring NMS, while the one-to-many head follows the traditional YOLO setting with NMS post-processing [18, 21]. In this thesis, both NMS-based and NMS-free detection settings are considered when analysing wake detection performance.

2.1.4 SAHI

Slicing Aided Hyper Inference (SAHI) is an inference strategy for improving small object detection in high-resolution images. It divides the original image into overlapping slices, applies the detector to each slice, and then merges the predictions back into the original image coordinate system [22]. Sliced inference allows the detector to process local regions at a higher effective scale, which may improve the detection of weak patterns. However, overlapping slices may also produce duplicate predictions, so prediction merging or NMS is required to obtain the final bounding boxes.



SAHI
Size = 640 * 640
Overlap rate = 0.1

SAHI
Size = 1024 * 1024
Overlap rate = 0.25

Figure 2.5: Illustration of SAHI in wake detection. The setup on the left side enabled the model to detect more vessels compared with the setup on the right side.

2.1.5 Detection Evaluation Metrics

The performance of object detection models was evaluated using standard object detection metrics, including precision, recall, average precision (AP), mean average precision at an IoU threshold of 0.50 (mAP50), and mean average precision averaged across multiple IoU thresholds (mAP50–95).

Precision measures the proportion of predicted detections that correspond to true wake instances, while recall measures the proportion of annotated wakes successfully detected by the model. They are defined as

$$Precision = \frac{TP}{TP + FP}, \tag{2.5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{2.6}$$

where TP denotes true positives, FP denotes false positives, and FN denotes false negatives.

To determine whether a prediction is correct, the Intersection over Union (IoU) between the predicted bounding box and the ground-truth annotation is computed:

$$IoU = \frac{Area(B_p \cap B_g)}{Area(B_p \cup B_g)}, \tag{2.7}$$

where B_p and B_g denote the predicted and ground-truth bounding boxes, respectively.

Average Precision (AP) summarizes the precision–recall relationship over all confidence thresholds and can be expressed as

$$AP = \int_0^1 P(R)\, dR, \tag{2.8}$$

where $P(R)$ represents precision as a function of recall.

For a multi-class detector, the mean Average Precision (mAP) is defined as

$$mAP = \frac{1}{N_c} \sum_{c=1}^{N_c} AP_c, \tag{2.9}$$

where N_c is the number of object classes.

Since wake detection is formulated as a single-class detection problem in this thesis, the mAP value is equivalent to the AP of the wake class.

The mAP50 metric evaluates AP at an IoU threshold of 0.50:

$$mAP50 = AP_{IoU=0.50}, \tag{2.10}$$

whereas mAP50–95 follows the COCO evaluation protocol and averages AP values over IoU thresholds ranging from 0.50 to 0.95:

$$mAP50\text{–}95 = \frac{1}{10} \sum_{t \in \{0.50, 0.55, \ldots, 0.95\}} AP_{IoU=t}. \tag{2.11}$$

These metrics jointly evaluate both detection completeness and localization accuracy and are therefore widely adopted in modern object detection research.

2.2 Image Preprocessing

Image processing is used at several stages of this thesis, from satellite image preparation to model input generation and speed estimation. This section introduces the main image processing methods used in the proposed workflow.

2.2.1 Noise Reduction

Noise reduction is applied to grayscale wake image crops before feature enhancement and speed estimation. In this thesis, four common filters were compared: mean filtering, Gaussian filtering, median filtering, and bilateral filtering. The purpose was to reduce background noise while preserving the main wake structure.

Mean filtering and Gaussian filtering smooth the image by averaging neighbouring pixels, but they may blur thin wake boundaries. Median filtering is effective for removing isolated bright or dark pixels, but it may also remove weak wake details in small image crops. Bilateral filtering considers both spatial distance and intensity similarity, allowing it to smooth homogeneous regions while preserving edges and linear structures. It can be expressed as

$$I_{\mathrm{bf}}(p) = \frac{1}{W_p} \sum_{q \in \Omega} I(q) \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2}\right) \exp\left(-\frac{(I(p)-I(q))^2}{2\sigma_r^2}\right), \tag{2.12}$$

where $I(p)$ is the intensity of the centre pixel, $I(q)$ is the intensity of a neighbouring pixel, Ω is the local filtering window, σ_s controls the spatial smoothing range, σ_r controls the influence of intensity differences, and W_p is the normalization factor.

For the wake image crops used in this thesis, bilateral filtering was selected as the final denoising method. It provides a suitable balance between noise reduction and wake-structure preservation before applying top-hat feature enhancement.

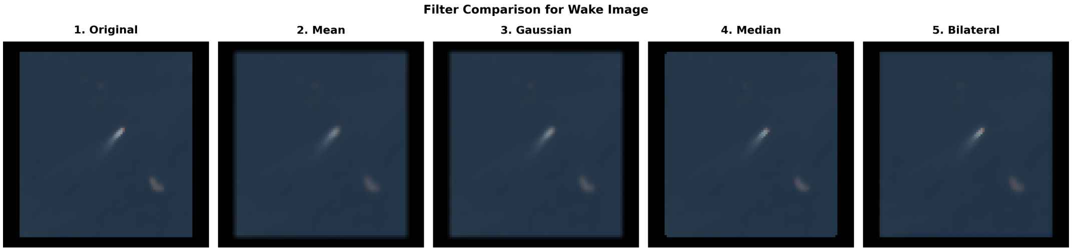


Figure 2.6: Comparison of denoising filters applied to a wake image crop. Bilateral filter displayed the best result in preserving wake structure and reducing background noise.

2.2.2 Feature Enhancement

Feature enhancement is used to improve the visibility of wake structures in the processed image crops. In this thesis, the white top-hat transform is applied after grayscale conversion and bilateral filtering. Since vessel wakes in Sentinel-2 imagery often appear as thin bright structures, this method can enhance wake-like patterns while suppressing slowly varying background intensity.

The white top-hat transform is defined as the difference between the input image and its morphological opening [23, 24]:

$$T_{\text{white}}(I) = I - (I \circ B), \tag{2.13}$$

where I is the input grayscale image, B is the structuring element, and $\circ$ denotes morphological opening. The opening operation is given by

$$I \circ B = (I \ominus B) \oplus B, \tag{2.14}$$

where $\ominus$ denotes erosion and $\oplus$ denotes dilation. The enhanced image is then used as input for the following speed estimation step.

2.3 Wake Morphology and Speed Dependence

When a vessel moves through water, momentum is transferred from the hull and propulsion system to the surrounding fluid, generating a visible wake pattern. A typical ship wake may include turbulent propeller wash, divergent waves, and transverse waves. In optical satellite imagery, these structures often appear as elongated bright or dark regions behind the vessel due to foam, surface roughness variations, and changes in reflected radiance [25]. Since wake formation is directly related to vessel motion, wake morphology can provide useful information for indirect speed estimation.

A key parameter describing wake formation is the Froude number:

$$Fr = \frac{v}{\sqrt{gL}} \tag{2.15}$$

where v is the vessel velocity, g is gravitational acceleration, and L is a characteristic vessel length. The Froude number describes the ratio between inertial and gravitational effects. As vessel speed increases, stronger wave systems and larger turbulent disturbances are usually produced, making the wake more visible in satellite imagery.

Ship-generated surface waves are often described by the Kelvin wake pattern. For a vessel moving in deep water at constant speed, the wake forms a characteristic V-shaped structure with a theoretical half-angle of approximately 19.47° [26]. Although this angle is independent of speed, the spatial scale of the wake varies with vessel speed and vessel size [5]. The characteristic wavelength of transverse waves can be expressed as:

$$\lambda = \frac{2\pi V^2}{g} \tag{2.16}$$

where V is the vessel speed and $g = 9.81\, m/s^2$. This equation indicates that wake wavelength increases quadratically with speed. Therefore, faster vessels tend to generate longer and more prominent wake signatures.

However, satellite sensors do not directly measure hydrodynamic wave fields. Wakes are observed through radiometric patterns caused by foam, turbulence, surface roughness changes, and reflection effects. At the 10 m spatial resolution of Sentinel-2 true-colour imagery, weak wakes may be difficult to distinguish from background sea clutter, especially for slow vessels or complex environmental conditions.

The relationship between wake appearance and vessel speed is also affected by wind, waves, illumination, viewing geometry, hull design, displacement, and propulsion system. Therefore, direct analytical speed estimation from satellite-observed wakes is challenging. In this thesis, the problem is formulated as an image-based regression task:

$$f(I) \rightarrow v \tag{2.17}$$

where I denotes the wake image and v denotes vessel speed. Instead of manually extracting hydrodynamic parameters, a convolutional neural network is used to learn visual features related to wake length, texture, turbulence intensity, and spatial extent directly from Sentinel-2 wake imagery.

2.4 Deep Learning-based Regression

The hydrodynamic characteristics discussed in section 2.3 indicate that vessel speed is related to observable wake properties, such as wake length, turbulence intensity, and overall morphology. However, analytical speed estimation from Sentinel-2 imagery is challenging because wake visibility is affected by image resolution, vessel characteristics, environmental conditions, and viewing geometry.

Therefore, this thesis adopts a data-driven approach to learn the relationship between wake appearance and vessel speed from AIS-labelled satellite imagery. Instead of manually defining wake descriptors, convolutional neural networks are used to automatically extract visual features and learn the non-linear mapping between wake patterns and vessel speed.

2.4.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) learn hierarchical spatial feature representations through successive layers of learned filters, pooling operations, and non-linear activations [27]. For image regression tasks where the output is a continuous scalar rather than a discrete class, the final classification head is replaced by a fully connected regression head with a single linear output neuron [28].

A convolution operation applies a learnable kernel K over an input image X:

$$Y(i,j) = \sum_m \sum_n X(i+m, j+n) K(m,n) \tag{2.18}$$

where Y is the output feature map, (i, j) is the pixel coordinate of X, and (m, n) is the pixel coordinate of K.

Through stacked convolutional layers, the network progressively learns hierarchical representations of low-level edges and brightness gradients, mid-level wake textures, and high-level wake morphology and turbulence structures [28].

The proposed CNN architecture consists of repeated convolutional blocks containing convolution, batch normalization, ReLU activation, and max pooling.

Pooling operations reduce spatial resolution while preserving dominant features:

$$Y(i, j) = \max_{(m,n)\in K} X(i + m, j + n) \tag{2.19}$$

This improves translational robustness and reduces computational complexity.

Batch normalization is a operation that stabilizes feature distributions during training:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{2.20}$$

where μ is the mean of x, σ^2 is the variance of x, and ϵ is a small constant added for numerical stability. This accelerates convergence and improves training stability.

The final regression head predicts continuous vessel speed values rather than discrete classes. Therefore, the task is formulated as a supervised regression problem:

$$\hat{v} = f_\theta(I) \tag{2.21}$$

where $\hat{v}$ is the speed estimation, I denotes the wake image, M denotes auxiliary metadata, f is the regression function and θ represents learnable CNN parameters.

2.4.2 Weighted Sampling for Data Imabalance

A common challenge in supervised vessel-speed regression is the highly imbalanced distribution of target values. In many maritime datasets, observations are concentrated within a limited range of moderate vessel speeds, whereas very slow and high-speed vessels are comparatively rare. Such imbalance can bias the optimization process toward densely populated regions of the target space, causing the model to achieve low overall error while performing poorly on underrepresented speed ranges.

From a statistical learning perspective, the empirical training loss is dominated by regions containing large numbers of samples. Consequently, the learned regression function tends to prioritize prediction accuracy around the most frequent target values. This phenomenon is particularly problematic when the objective is to achieve reliable performance across the full operational speed range rather than only the most common vessel speeds.

A common strategy for mitigating this issue is importance-weighted sampling:

$$\omega_i = \frac{1}{N_{b_i}} \tag{2.22}$$

where ω_i is the sampling probability of a sample i, b_i is the speed bin sample i belongs to, and N_{b_i} is the number of samples in the corresponding speed bin.

This method allows samples from sparsely populated regions receive larger weights than samples from densely populated regions. During stochastic optimization, this increases the probability that rare samples contribute to parameter updates, leading to a more balanced representation of the target distribution within training batches.

For vessel-speed estimation, weighted sampling is particularly beneficial because high-speed vessels are typically underrepresented in AIS-derived datasets. By increasing their effective contribution during optimization, the learning process becomes less biased toward moderate-speed vessels and can achieve improved generalization across the entire speed range.

To implement weighted sampling, the vessel-speed distribution was divided into five predefined intervals:

$$\mathcal{B} = \{B_1, B_2, B_3, B_4, B_5\} = [0,3), [3,7), [7,12), [12,25), [25,40) \text{ kn} \tag{2.23}$$

2.4.3 Huber Regression Loss

Wake-speed labels derived from AIS matching contain unavoidable uncertainty due to AIS temporal mismatch, geolocation inaccuracies, vessel maneuvering, environmental effects, imperfect wake visibility.

To improve robustness against noisy labels and outliers, Huber loss [29] was chosen which combines the advantages of L1 and L2 regression:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2, & |a| \le \delta \\ \delta(|a| - \frac{1}{2}\delta), & |a| > \delta \end{cases} \tag{2.24}$$

where a is the residual, and δ is a hyperparameter where A smaller δ makes the loss closer to Mean Absolute Error (MAE), while a larger δ makes the loss closer to Mean Absolute Error (MSE). For small prediction errors, the loss behaves quadratically like MSE, enabling stable optimization. For large errors, it becomes linear like MAE, reducing sensitivity to mislabeled samples and extreme outliers.

This property ensures the robustness for satellite-derived AIS datasets where exact synchronization between wake appearance and AIS-reported speed cannot always be guaranteed.

2.4.4 Regression Evaluation Metrics

The performance of the vessel-speed regression model was evaluated using standard regression metrics, namely Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Both metrics quantify the discrepancy between predicted vessel speeds and the corresponding ground-truth AIS measurements.

The Mean Absolute Error is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|, \tag{2.25}$$

where y_i and $\hat{y}_i$ denote the ground-truth and predicted vessel speeds, respectively, and N is the total number of samples.

MAE measures the average magnitude of prediction errors and is expressed in the same physical unit as the target variable (knots). Since all errors contribute linearly, MAE provides an intuitive interpretation of the average prediction deviation.

The Root Mean Square Error is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}, \tag{2.26}$$

which assigns a larger penalty to large prediction errors due to the quadratic term. Consequently, RMSE is more sensitive to outliers than MAE and provides additional information regarding prediction robustness.

Because the wake–speed dataset exhibits an imbalanced target distribution, overall metrics alone may not adequately describe model behaviour across the entire speed range. Therefore, segmented evaluation was also performed by computing MAE within predefined speed intervals. This enables the analysis of model performance for both common and underrepresented vessel-speed ranges.

To assess whether the regression model learns meaningful wake–speed relationships beyond simple statistical priors, a mean baseline predictor was introduced. The baseline predicts a constant speed equal to the average speed observed in the training dataset:

$$\hat{y}_{base} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{2.27}$$

The relative improvement of the proposed model over the baseline can then be quantified as:

$$Improvement = \frac{MAE_{baseline} - MAE_{CNN}}{MAE_{baseline}} * 100\% \tag{2.28}$$

This comparison provides an intuitive measure of the practical value of the learned regression model. A large improvement indicates that the model successfully extracts speed-related information from wake morphology rather than relying solely on the average speed of the training dataset.

2.5 Underwater Noise Model

2.5.1 Sources of Vessel Noise

Underwater radiated noise (URN) is the sound emitted by vessels into the surrounding aquatic environment. For motorised recreational boats, the dominant noise sources include engine operation, gearbox vibrations, propeller rotation, cavitation, and hydrodynamic interactions between the hull and the water column [8]. These mechanisms generate broadband acoustic energy that can propagate over considerable distances and contribute to underwater sound pollution.

The magnitude of vessel-generated noise depends on multiple factors, including vessel size, propulsion system, operating condition, and vessel speed. In particular, vessel speed plays a central role because increasing speed generally leads to stronger propeller loading, higher cavitation probability, increased turbulence, and more energetic wake generation. Consequently, vessel speed is commonly regarded as one of the primary predictors of underwater radiated noise levels.

The increasing concern regarding anthropogenic underwater noise has motivated the development of spatial noise assessment methods capable of identifying areas exposed to elevated vessel activity. Such methods are particularly relevant in coastal waters where recreational boating activity is intensive and AIS coverage is often incomplete.

2.5.2 Speed-Noise Relationship

Several studies have demonstrated a positive correlation between vessel speed and underwater radiated noise levels [30]. Although the exact relationship depends on vessel characteristics and operating conditions, source levels generally increase with speed due to stronger propulsion-induced noise and increased cavitation activity.

To represent this relationship, a simplified speed-dependent source-level model was adopted in this thesis:

$$SL_i = SL_{ref} + \alpha \log_{10}\left(\frac{V_i}{V_{ref}}\right), \tag{2.29}$$

where SL_i is the estimated source level indicator of vessel i, V_i is the vessel speed in knots, V_{ref} is a reference speed, SL_{ref} is the reference source level at V_{ref}, and α controls the speed dependence.

Equation 2.29 represents the simplest speed-based source-level indicator and assumes that acoustic emission increases logarithmically with vessel speed [30]. Although simplified, the model provides a practical mechanism for linking speed estimation results to underwater noise generation and serves as the baseline source-level model used throughout this thesis.

2.5.3 Acoustic Propagation Loss

After acoustic energy is generated at the source, the sound propagates through the marine environment and gradually attenuates with distance. The attenuation process is commonly described through transmission loss (TL), which accounts for geometric spreading and medium absorption.

In this thesis, a simplified free-field propagation model was employed:

$$TL(r) = 20 \log_{10}(r) + \beta r \tag{2.30}$$

where r denotes source–receiver distance in meters and β represents shallow-water absorption loss.

The first term models spherical spreading of acoustic energy, while the second term approximates attenuation caused by absorption and scattering processes within the water column. Although more advanced propagation models exist, the simplified formulation is sufficient for demonstrating the integration of satellite-derived vessel information into a spatial noise-mapping framework.

2.5.4 Noise Mapping

Noise maps are widely used to visualize the spatial distribution of underwater acoustic energy. Existing large-scale maritime noise-monitoring systems commonly combine AIS vessel tracks, source-level models, and acoustic propagation models to estimate vessel contributions across large geographic areas [1].

In this thesis, vessel speeds estimated from satellite-observed wake patterns are used as an alternative source of information for constructing a simplified spatial noise indicator. The objective is not to produce a fully calibrated acoustic noise map, but rather to demonstrate the feasibility of integrating satellite-based wake detection and speed estimation into a vessel-noise monitoring workflow.

The received sound pressure level at each spatial location was computed by incoherently summing acoustic energy contributions from all vessels:

$$SPL(x, y) = 10 \log_{10} \left(\sum_i 10^{(SL_i - TL_i(x,y))/10} \right) \tag{2.31}$$

where SL_i denotes the source-level indicator of vessel i and $TL_i(x, y)$ represents the corresponding transmission loss between the vessel and the receiver location.

The resulting spatial distribution forms a relative noise map that highlights regions potentially exposed to elevated vessel-generated underwater noise. Although the framework relies on simplified acoustic assumptions, it provides a proof-of-concept demonstration of how satellite remote sensing, vessel wake analysis, and deep learning can be integrated for environmental monitoring applications.

3

Data Sources

This chapter presents the data sources used in this thesis. It describes the Sentinel satellite imagery and AIS data used for wake detection, vessel matching, and speed estimation. It also introduces the data selection process and key limitations of the datasets.

3.1 Sentinel-2 Optical Imagery

Before constructing the wake detection dataset, several satellite image sources were compared over the same study area. The purpose of this comparison was to identify an image source that is suitable for detecting small recreational-vessel wakes in coastal waters. Four image sources were considered: Sentinel-1, Sentinel-2, Sentinel-3, and Landsat 8/9. These sources differ in sensor type, spatial resolution, visual interpretability, and suitability for wake detection.

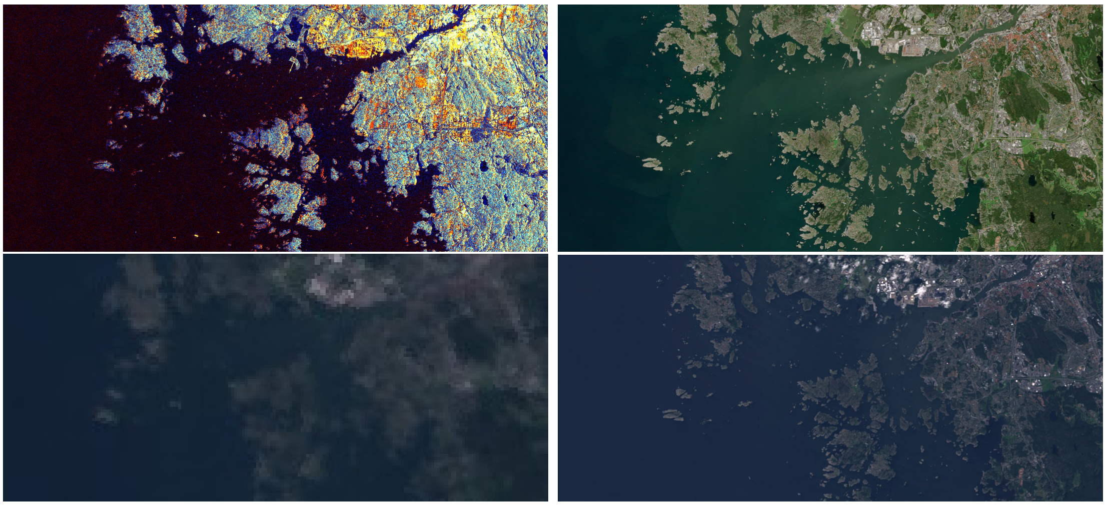


Figure 3.1: Satellite images from different sources over the same study area: Sentinel-1, Sentinel-2, Sentinel-3, and Landsat 8/9

Figure 3.1 shows examples from the same coastal area using different satellite image sources. Sentinel-1 provides synthetic aperture radar (SAR) imagery [31]. It has the advantage of all-weather and day-night imaging capability, however, the image is not a true colour optical image, and wake patterns appear as radar backscatter structures rather than visually interpretable optical features.

Sentinel-2 provides multispectral optical imagery, including true colour images at 10 m spatial resolution. In the Sentinel-2 example, coastal structures, islands, sea surface patterns, vessels, and visible wakes can be interpreted clearly [32]. Although the 10 m resolution is still coarse for very small vessels, it provides a useful balance between spatial detail and large-area coverage. By comparing visible wake structures with AIS speed records, it was observed that some wakes from vessels travelling below 5 knots could still be identified in Sentinel-2 imagery. This suggests that the 10 m resolution is sufficient for detecting many visible recreational-vessel wakes in the selected coastal scenes. Therefore, Sentinel-2 is suitable for constructing a coastal wake detection dataset over the Gothenburg archipelago.

Sentinel-3 and Landsat 8/9 also provide optical imagery and have wide-area coverage, but their spatial resolution are much coarser [33] [34]. In the example images, the study area appears blurred, and small islands, narrow waterways, vessels, and wakes are not clearly visible.

The comparison is summarized in Table 3.1. Based on this comparison, Sentinel-2 was selected as the primary image source for this thesis. It provides true colour optical imagery, 10 m spatial resolution, wide spatial coverage, frequent revisit, and open access. These properties make it the most suitable option among the compared data sources for detecting visible recreational-vessel wakes in large coastal scenes.

Source	Type	Resolution	Wake visibility	Suitability
Sentinel-1	SAR	down to 5 m	Different SAR pattern	Limited
Sentinel-2	Optical	10 m	Clear in many scenes	High
Sentinel-3	Optical	300 m	Too coarse	Low
Landsat 8/9	Optical	30 m / 15 m PAN	Weak or unclear	Medium-low

Table 3.1: Comparison of satellite imagery sources for wake detection

3.1.1 Sentinel-2 Tile Metadata

The acquisition time of each Sentinel-2 image was obtained from the tile-level metadata file, MTD_TL.xml. This file contains detailed information about the Sentinel-2 tile, including the sensing time, tile identifier, coordinate reference system, spatial resolution, and image dimensions. In this thesis, the sensing time was used as the satellite image timestamp for matching detected wake positions with AIS records.

For each Sentinel-2 scene, the relevant sensing time was extracted from the metadata and converted to UTC. This timestamp was then used as the centre of the AIS matching time window.

Table 3.2: Sentinel-2 tile metadata extracted from the MTD_TL file for the 2024-09-08 scene.

Metadata field	Value
Satellite mission	Sentinel-2B
Processing level	Level-2A
L1C tile ID	L1C_TL_20240908T131436_A039216_T32VPJ
L2A tile ID	L2A_TL_20240908T135227_A039216_T32VPJ
Datastrip ID	L2A_DS_20240908T135227_S20240908T103625
Sensing time	2024-09-08T10:44:58.788712Z
Coordinate reference system	WGS84 / UTM zone 32N
EPSG code	EPSG:32632
Tile ID	T32VPJ
Absolute orbit number	039216
Processing baseline	N05.11

Listing 3.1 presents the relevant metadata fields from the Sentinel-2 MTD_TL file. The `SENSING_TIME` value was used as the image acquisition time and served as the reference timestamp for matching detected wakes with AIS records.

Listing 3.1: Excerpt from the Sentinel-2 MTD_TL metadata file for the 2024-09-08 scene.

```
<n1:General_Info>
  <L1C_TILE_ID metadataLevel="Brief">
    S2B_OPER_MSI_L1C_TL_2BPS_20240908T131436_A039216_T32VPJ_N05.11
  </L1C_TILE_ID>

  <TILE_ID metadataLevel="Brief">
    S2B_OPER_MSI_L2A_TL_2BPS_20240908T135227_A039216_T32VPJ_N05.11
  </TILE_ID>

  <DATASTRIP_ID metadataLevel="Standard">
    S2B_OPER_MSI_L2A_DS_2BPS_20240908T135227_S20240908T103625_N05.11
  </DATASTRIP_ID>

  <SENSING_TIME metadataLevel="Standard">
    2024-09-08T10:44:58.788712Z
  </SENSING_TIME>

  <Archiving_Info metadataLevel="Expertise">
    <ARCHIVING_CENTRE>2BPS</ARCHIVING_CENTRE>
    <ARCHIVING_TIME>2024-09-08T15:01:49.235698Z</ARCHIVING_TIME>
  </Archiving_Info>
</n1:General_Info>
```

3.2 AIS Data

The vessel data used in this study were obtained from Automatic Identification System (AIS) records. AIS data provide vessel traffic information, including vessel position, timestamp, speed, course, vessel identity, and vessel type. In this thesis, AIS records were mainly used for two purposes. First, they were used as a reference source to compare with the vessel and wake detections obtained from satellite images. Second, they were used to generate speed labels for detected wakes. For this purpose, wake detections from Sentinel-2 images were spatially and temporally matched with AIS records using geolocation and imaging timestamps. The matched samples were then used to build a speed–wake dataset for speed estimation.

To further understand the characteristics of the available AIS data, a statistical analysis was conducted using AIS records from the Baltic Sea region on 2024-09-08. The dataset includes all available vessel types in the selected spatial and temporal range. This analysis was used to examine the vessel type distribution, speed distribution, spatial distribution, and the representation of type 37 vessels, which are relevant to recreational boating activities. Table 3.3 summarizes the main AIS fields used in this thesis.

Item	Value
Date	2024-09-08
Region	Baltic Sea region
Geographic ranvge	10.6–20.0°E, 55.2–60.2°N
Raw AIS records	4,278,060
Valid AIS records after cleaning	4,258,764
Records after geographic filtering	2,058,356
Vessels after geographic filtering	6,381
Records in selected time window	1,370,161
Vessels in selected time window	6,194
Main variables used	Position, timestamp, SOG, MMSI, ship type

Table 3.3: AIS data used for the statistical analysis

Figure 3.2 compares the speed distribution of the 15 most frequent AIS vessel types in the selected dataset, the red box indicates recreational vessels. The figure shows that the dominant vessel types have different speed characteristics. Some categories contain many stationary or very slow records, while others show higher median speeds and wider speed ranges. For recreational vessels, the speed distribution includes both low-speed records and moving records that may be relevant for wake generation.

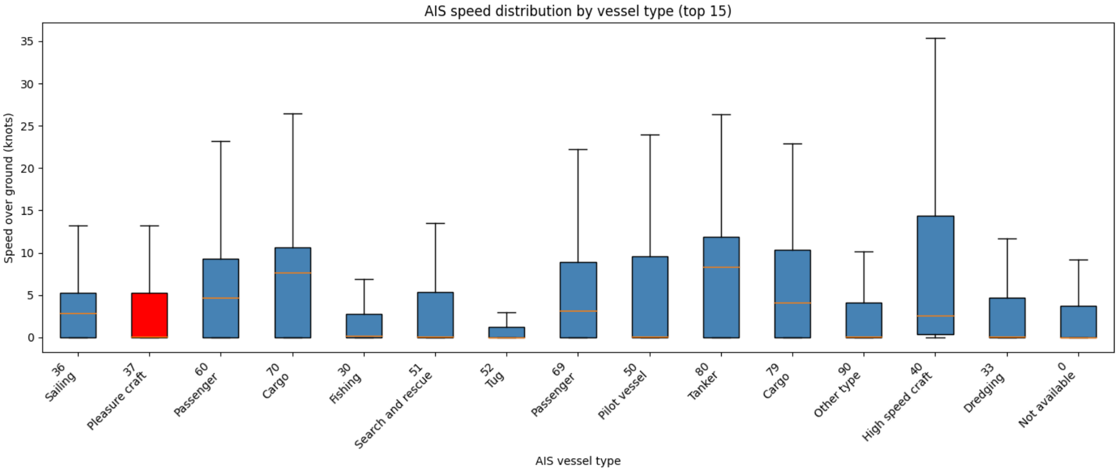


Figure 3.2: Speed distribution of the 15 most frequent AIS vessel types

Figure 3.3 shows the speed distribution of recreational vessels in the selected AIS dataset. Most records are below 10 knots, with large numbers in the 0–0.5 knot, 0.5–5 knot, and 5–10 knot ranges. Records above 10 knots are less frequent, while speeds above 30 knots account for only a small proportion of the data. This indicates that the AIS records for recreational vessels cover a wide speed range, but are mainly concentrated at low to moderate speeds.

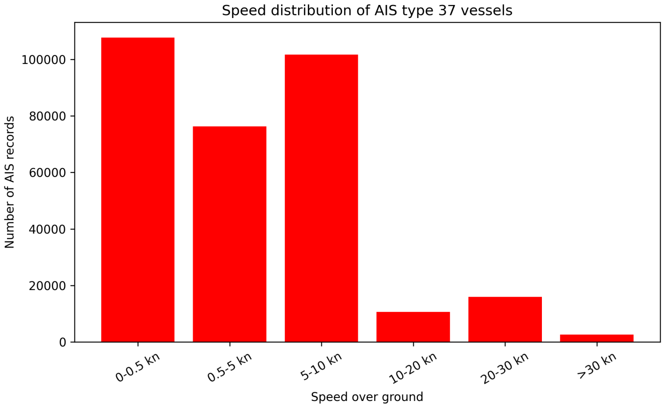


Figure 3.3: Speed distribution of recreational vessels

Figure 3.4 shows the number of unique AIS vessels by vessel type in the selected dataset. AIS type 37, which is treated as recreational vessels in this thesis, is the most frequent vessel type in the AIS data. However, the number of AIS-visible recreational vessels is still much smaller than the estimated total number of recreational boats in Sweden. Previous studies report approximately 800,000–900,000 recreational boats in Sweden [35], and about 860,000 leisure boats in Sweden [36]. This difference suggests that AIS data capture only part of the recreational-boat activity, since many small vessels are not represented in AIS records.

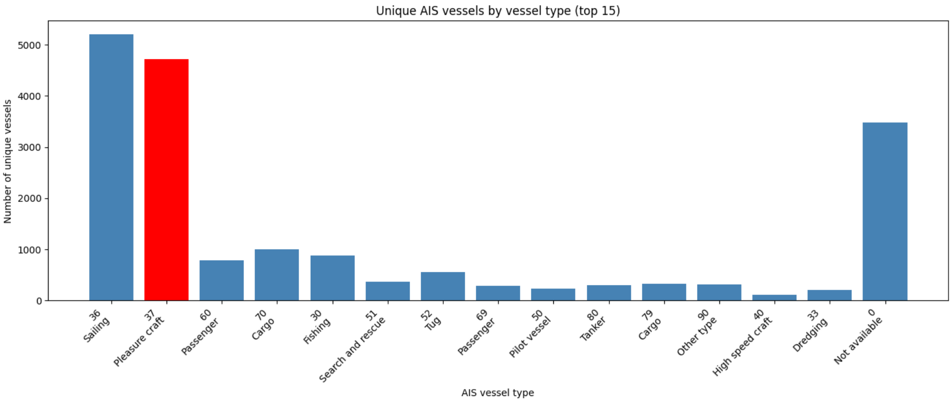


Figure 3.4: AIS type distribution

Figure 3.5 further illustrates this limitation. In the selected Sentinel-2 image of the southern Gothenburg archipelago, only 11 AIS vessel records are visible, while more wake structures can be observed visually in the image. This shows that AIS data alone cannot fully describe small-vessel activity in coastal areas. Satellite imagery is therefore used as a complementary data source for detecting visible wakes.

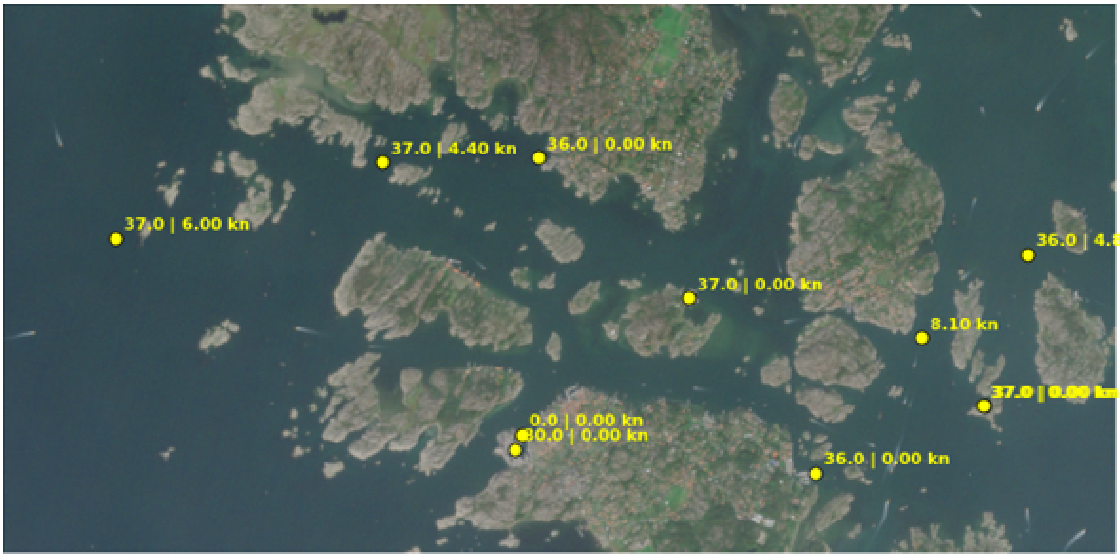


Figure 3.5: AIS records overlaid on satellite imagery

4

Methods

This chapter describes the design and implementation of the workflow proposed in this thesis. It presents the dataset construction, AIS-satellite matching, wake detection, speed estimation, and simplified noise mapping methods, followed by the evaluation procedures used for each step.

4.1 Overall Workflow

The proposed AI-based workflow, which is illustrated in Figure 4.1, is designed as a two-stage machine learning pipeline. The first stage detects visible wake patterns from Sentinel-2 satellite images using a YOLO-based object detection model. The input is a satellite image or image tile, and the output is a set of bounding boxes representing possible vessel wakes. These detections are then cropped from the original image and processed as wake image samples.

The second stage estimates vessel speed from the detected wake crops. Before speed estimation, the wake crops are denoised and enhanced to make the wake structures more visible. A CNN-based regression model is then used to predict vessel speed from the processed wake images. The estimated speeds are further used as input for the simplified underwater noise mapping case study, where the detected wake locations and predicted speeds are converted into a spatial noise indicator for the selected coastal area.

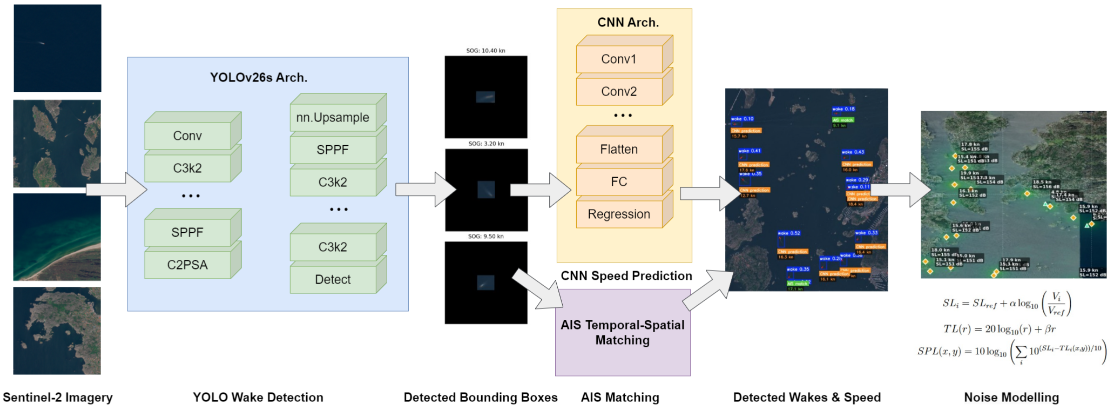


Figure 4.1: Overall workflow. The workflow goes from Sentinel-2 imagery, then to YOLO detection and AIS matching/CNN speed prediction, and finally to noise modelling.

Overall, the proposed workflow provides an end-to-end AI solution for vessel activity monitoring and environmental noise study.

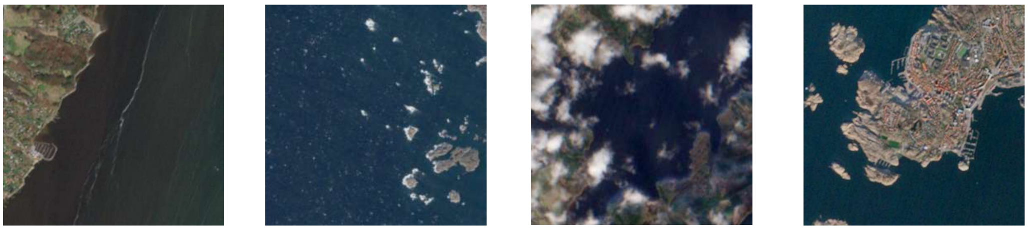
4.2 Wake Dataset from Sentinel-2 Imagery

4.2.1 Dataset Construction

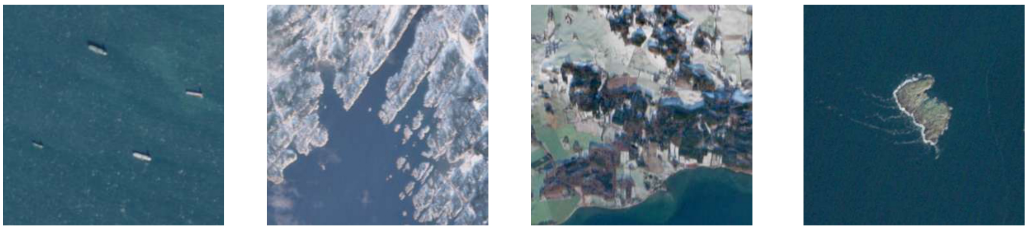
The wake detection dataset was constructed from Sentinel-2 true colour imagery covering Swedish coastal areas. The selected images were mainly collected from different periods in 2024. The annotation target in this dataset is the visible vessel wake rather than the vessel hull. Each wake was annotated using a horizontal bounding box (HBB). As shown in figure 4.2.

Figure 4.2: Wake dataset examples

Hard negative samples were included to improve the robustness of the detection model. These samples contain image regions that may visually resemble wakes but do not correspond to actual vessel wakes. By including these difficult background cases, the dataset encourages the model to distinguish wake structures from similar non-wake patterns.



surface wave streaks　skerries / sea clutter　clouds　port facilities



ships without wakes　snow-covered coast　bright urban structures　bright coastline

Figure 4.3: Hard negative samples examples

The dataset also includes repeated or similar coastal areas observed under different seasonal and imaging conditions. This design allows the trained model to be evaluated under more realistic coastal remote sensing conditions, where wake visibility is affected by background complexity, cloud cover, illumination, and sea-surface state.

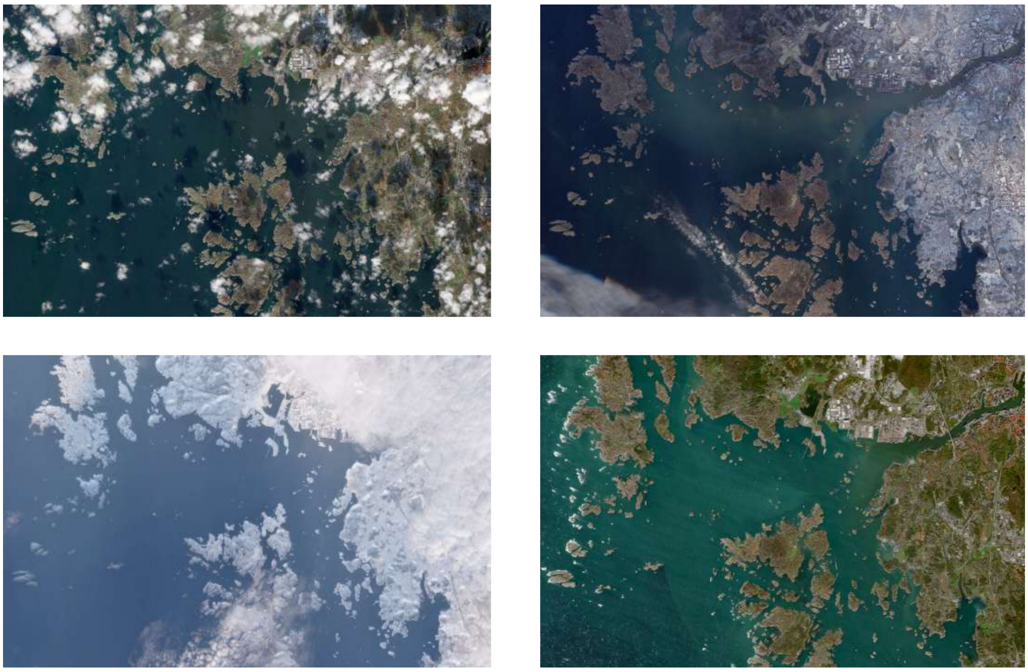


Figure 4.4: Same area in different season and conditions

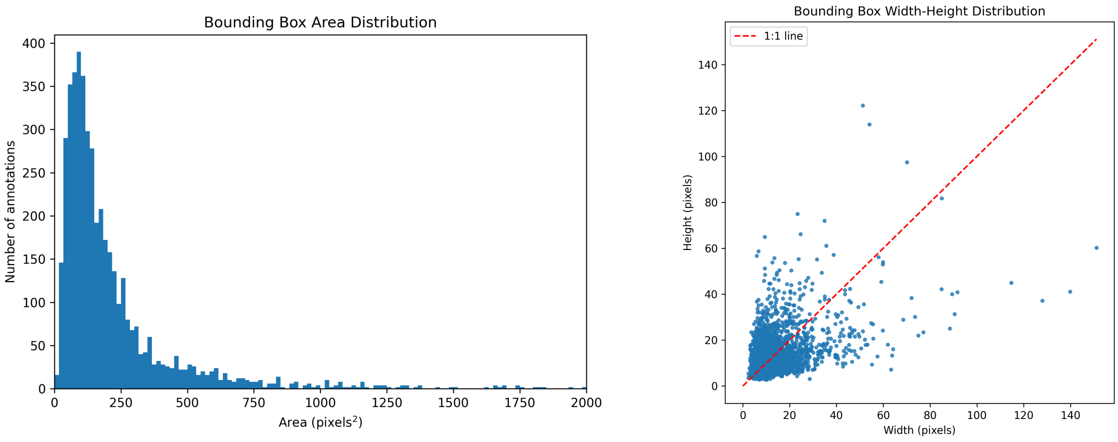
4.2.2 Dataset Statistics

The final Sentinel-2 wake detection dataset consists of 256 × 256 pixel image tiles annotated in YOLO horizontal bounding box format. The dataset includes both wake-containing images and negative samples, and is split into training and validation sets, as summarized in Table 4.1.

Item	Value
Total images	3734
Training images	2612
Validation images	1122
Images with wake labels	1612
Images without wake labels	2122
Total wake annotations	4640
Mean bounding box width	14.73 pixels
Mean bounding box height	14.48 pixels
Mean bounding box area	254.48 pixels^2
Median bounding box area	138.07 pixels^2

Table 4.1: Summary of the Sentinel-2 wake detection dataset

Figure 4.5a shows the bounding box size distributions of the Sentinel-2 wake detection dataset. The left subplot presents the bounding box area distribution, where most annotations are concentrated in the small-area range. The median bounding box area is 138.07 pixels^2, and the mean area is 254.48 pixels^2, indicating that most wake targets occupy only a small part of the 256 × 256 image tiles.



(a) Bounding box area distributions

(b) Bounding box width-height distributions

Figure 4.5: Sentinel-2 wake detection dataset bounding box statistics.

Figure 4.5b shows the width-height distribution of the annotated boxes. Most samples are located in the lower-left region, confirming that the dataset mainly contains small wake objects. The red dashed line represents the 1:1 reference line, and the spread of points away from this line shows that wake annotations have variable shapes, including elongated and partially visible wake structures.

4.3 Speed-Wake Dataset Construction

4.3.1 Speed-Wake Sample Generation Pipeline

A dedicated speed-wake regression dataset was constructed by combining YOLO wake detection bounding boxes, AIS records-bounding box vessel matching, and image preprocessing to enable the study of wake-based vessel speed estimation. Since no publicly available dataset exists for this task, the wake–speed dataset generation pipeline was designed and implemented specifically for this research. The pipeline consists of five stages: Sentinel-2 image acquisition, GeoTIFF tiling, YOLO-based vessel detection , AIS temporal-spatial matching, and wake-speed sample generation with data augmentation.

The complete pipeline from raw satellite imagery to vessel speed estimate operates in five stages, as illustrated in Figure 4.6. Each stage is automated and requires no manual intervention beyond initial configuration.

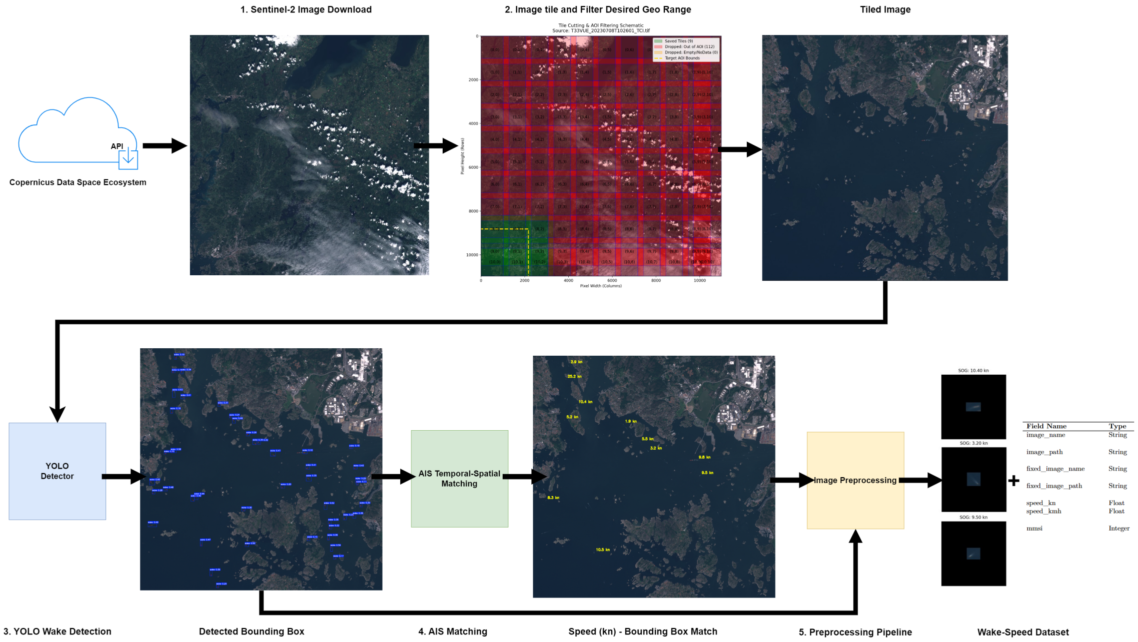


Figure 4.6: Speed-Wake Dataset Construction. The pipeline consists of Sentinel-2 images download, image tiling and geographical filtering, YOLO detection, AIS matching, and image preprocessing.

Sentinel-2 Level-2A products were queried and downloaded via the CDSE OData API. Products were filtered by: tile identifier (MGRS grid cell covering the study area in the Gothenburg archipelago), acquisition date range (June–August 2023 and

2024, summer recreational boating season), and cloud cover below 10%. Only images with sufficiently low cloud coverage were selected in order to minimize occlusion over maritime regions. The selection process was restricted to coastal and harbor regions with high recreational boat activity.

Each Sentinal-2 product archive was tiled into $1280 * 1280$ pixel patches using a sliding window with 256-pixel overlap (stride = 1024 pixels), retaining only tiles whose geographic extent intersects the defined area of interest. A preprocessing stage was applied to remove invalid image tiles dominated by empty or black regions. Tiles containing insufficient valid pixel information were discarded automatically. This significantly reduced unnecessary inference time and prevented false detections caused by image borders or missing data regions. The resulting image tiles formed the input for the wake detection stage. The process above is shown in Figure 4.7. Geographic coordinates were preserved using the rasterio library, allowing each pixel to be mapped back to WGS84 longitude and latitude via UTM affine transformation.

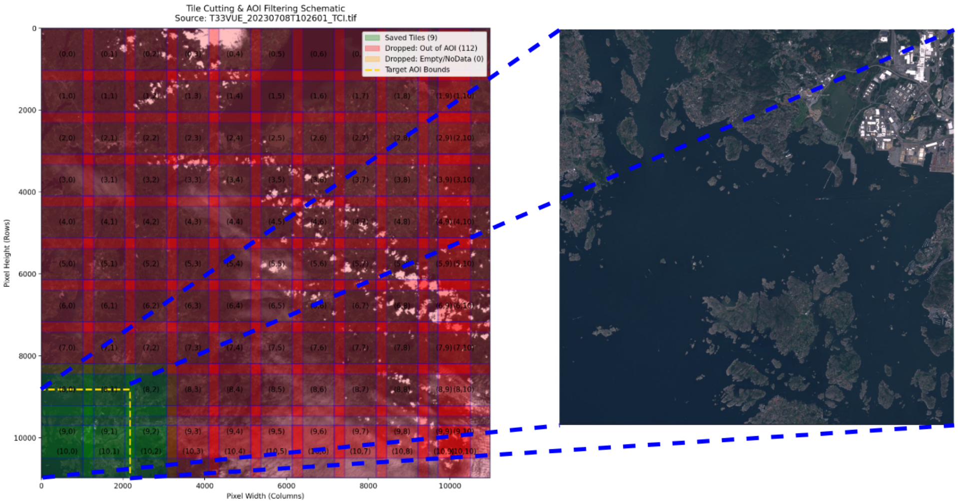


Figure 4.7: Sentinel-2 image tiling. The raw Sentinel-2 images was tiled into $1280 * 1280$ tiles. Only the tiles fall within the desired geographical area were kept (green), while the rest of them were discarded (red).

The downloaded images were stored as multi-band GeoTIFF files preserving geospatial metadata and coordinate reference system information for later stages.

Vessel detection was performed using the YOLOv26 model trained in the earlier stage wtih Sentinel-2 images from the study area. Inference was run at 1280-pixel input resolution with a confidence threshold of 0.10 and IoU threshold of 0.20, producing axis-aligned bounding boxes around each detected vessel.

For each wake detected by the YOLO model, the corresponding bounding box was used to crop the wake region from the original Sentinel-2 image. To estimate the vessel position inside each detected wake box, the ten brightest pixels within the bounding box were selected. Their average position was used as an approximate vessel location. This step was used because the vessel body is often brighter than the surrounding wake and water surface in true colour satellite imagery.

Each detected wake was matched with nearby AIS trajectories using spatial proximity constraints. For every wake detection, the geospatial coordinates of the localized bright point were transformed from image coordinates into geographic coordinates using the GeoTIFF spatial transform. AIS messages within the acquisition time window were retrieved. The estimated vessel position was then matched with AIS data using both temporal and spatial information. For each image, an AIS search window of one minute before and one minute after the satellite acquisition timestamp was created. All AIS records within this time window and within the selected geographic range were first extracted. Since Sentinel-2 imaging is not instantaneous and the actual wake formation time may differ from the nominal image timestamp by several seconds, the final matching was based on spatial proximity. For each detected vessel position, the nearest AIS record was selected as the matched vessel record. The closest valid AIS candidate was assigned as the corresponding vessel. To avoid repeated assignment of the same vessel, each MMSI number was allowed to appear only once in the final matched dataset.

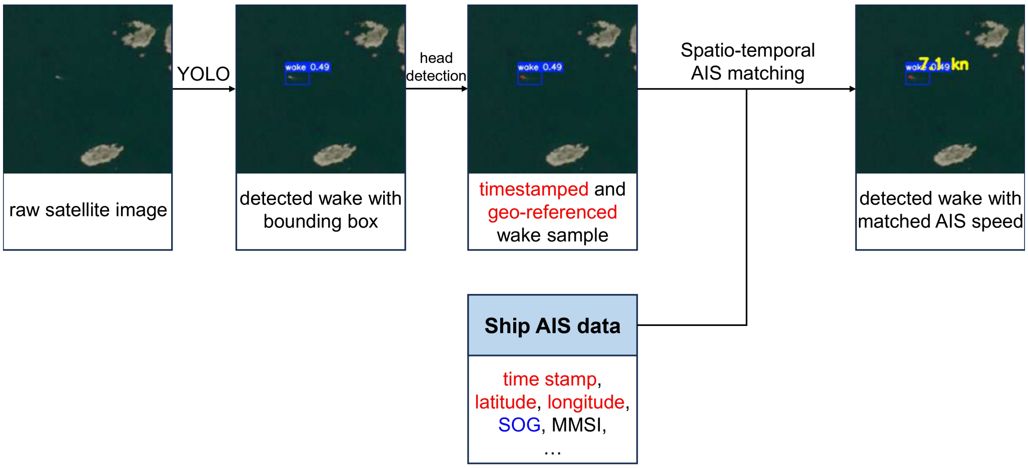


Figure 4.8: AIS matching workflow. The steps include YOLO detection, wake-head detection, and nearest neighbor mathtching.

The following metadata were stored together with the cropped bounding box: Vessel MMSI, Speed over ground (SOG), Matching distance, Geographic coordinates, Bounding box pixel coordinates.

To reduce label noise, detections with excessively large matching distances were filtered during dataset construction. Extremely low-speed (under 0.5 kn) vessels were also removed because weak or stationary wakes often produce ambiguous visual structures.

Setting	Value
Wake source	YOLO detection results
Vessel position estimation	Average position of the ten brightest pixels
Satellite timestamp	Sentinel-2 image sensing time
AIS temporal window	± 1 min
Spatial matching threshold	200 m
Image size	100 × 100 pixels
AIS spatial filtering	AIS records within the selected geographic range
Matching rule	Nearest AIS record to the estimated vessel position
Speed label	SOG

Table 4.2: Main settings used for AIS-satellite matching and speed-wake dataset construction.

After AIS matching, individual wake-centered image crops were extracted from the original GeoTIFF imagery. The initial YOLO bounding boxes were expanded using proportional padding in order to preserve surrounding wake structures.
This padding step was necessary because vessel wakes often extend significantly beyond the detected hull region. Crops that exceeded the target size or contained invalid dimensions were discarded to filter out the abnormal samples. Each crop was then converted into RGB format and normalized into 8-bit images. Since Sentinel-2 images may contain different reflectance ranges, adaptive intensity normalization was applied. Bilateral filtering was further used to suppress sensor noise while preserving wake edges and texture structures. These steps effectively reduced background noise and improve the visibility of wake structures.

All samples were finally padded to a fixed spatial resolution of $100 * 100$ pixels. Unlike direct resizing, zero-padding preserved the original wake geometry and aspect ratio without introducing interpolation distortion. The image preprocessing process is shown in Figure 4.9.

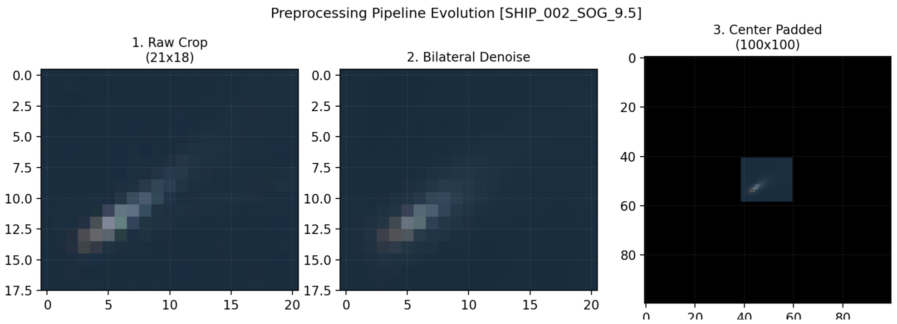


Figure 4.9: Wake image preprocessing pipeline. Bilateral filtering and image padding were used, which reduced background noise, improved the visibility of wake structures, and standardize the input size.

Upon this step, the following metadata were stored for each sample: Bounding-box coordinates, Crop dimensions, Vessel speed, Matching distance, Source image information, Bright pixel count, and CRS metadata.

The collected wake-speed dataset exhibits strong imbalance across speed ranges, particularly for high-speed (over 15 kn). As discussed in 3.2, the majority of available AIS records correspond to vessels travel at moderate speeds ($0.5 - 10$ kn), while records of high-speed (over 10 kn)vessels are comparatively rare. The same speed distribution pattern is found in the wake-speed dataset. Such imbalance biases the model toward dominant speed intervals. To mitigate this issue, weighted random sampling was introduced during training. Samples belonging to underrepresented speed bins received larger sampling probabilities. This strategy increases the effective contribution of rare high-speed samples during optimization and improves regression performance across the full speed range.

Besides the speed bias of samples, remote sensing wake datasets are typically limited in size and exhibit varying illumination conditions, noise levels, and observation geometries. To handle the unbalanced and limited dataset, data augmentation was therefore applied to improve generalization and reduce overfitting.

The augmentation pipeline is shown in Figure 4.10. Horizontal flipping preserves wake-speed relationships because left-right wake orientation does not fundamentally alter vessel velocity characteristics. Random noise injection simulates sensor variability and environmental disturbances. Gamma correction models illumination variation caused by atmospheric and solar reflection differences.

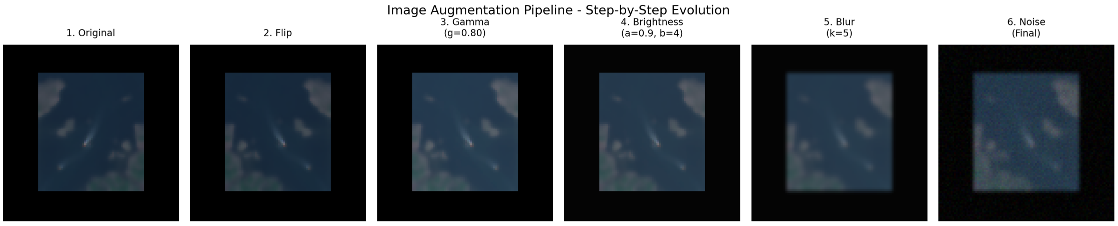


Figure 4.10: Augmentation workflow. The pipeline included horizontal flipping, Gaussian noise injection, and gamma correction.

The augmentation operations introduced more diverse samples. This has improved the usability of the dataset, ensured model robustness against variations in imaging conditions, and reduced overfitting, while preserving the underlying wake morphology relevant for speed estimation.

4.3.2 Statistics of the Constructed Dataset

After preprocessing and filtering, the resulting dataset consisted of paired wake images and vessel speed labels suitable for supervised regression. The dataset contains RGB wake crops, associated vessel speeds in knots, bounding-box metadata, and

AIS matching information.

The dataset exhibits a long-tailed speed distribution with a dominant concentration in medium-speed vessels and relatively few high-speed samples. This imbalance motivated the weighted sampling and stratified evaluation strategies described in the following sections.

The statistics of different speed bins in the generated dataset are summarized in Table 4.3.

Table 4.3: Wake-speed dataset statistics

Category	Count	Percentage
0–3 kn	29	3.3%
3–7 kn	126	14.4%
7–12 kn	527	60.2%
12–25 kn	155	17.7%
25–40 kn	38	4.3%
Total	875	100%

The dataset was divided into training, validation, and test subsets using stratified splitting. The final split ratios are shown in Table 4.4,.

Table 4.4: Dataset split statistics

Subset	Samples	Ratio	Purpose
Training	684	78.2%	Model optimization
Validation	95	10.9%	Hyperparameter tuning
Test	96	10.9%	Final evaluation

Finally, the augmentation operations were applied to high-speed vessels samples in the training set to address the issue of data unbalance in speed, where the threshold of lower speed was set as 20 kn. The comparison of the numbers of samples in different speed ranges is shown in Figure 4.11.

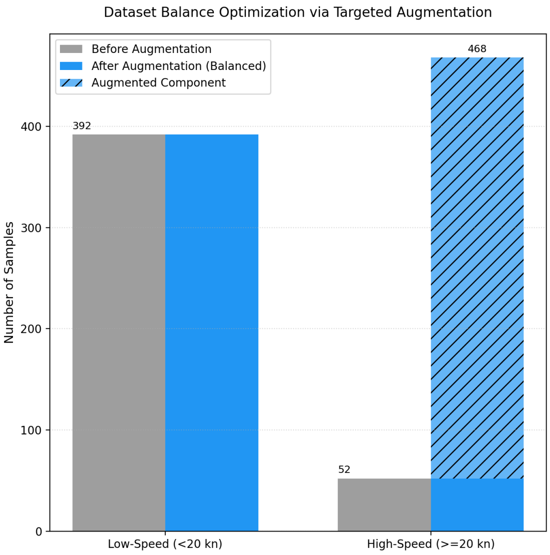


Figure 4.11: Data volume comparison after augmentation. The representation of high speed samples was substantially increased.

4.4 YOLO-based Wake Detection Framework

4.4.1 Model

A YOLO-based object detection model was used to detect visible wake patterns in Sentinel-2 satellite images. The detection task was formulated as a single-class object detection problem, where the target class is the vessel wake. In this study, YOLOv26s was used as the main detection model because of its improved capability for small-object detection. The model was initialized with official COCO-pretrained weights and then fine-tuned on the constructed Sentinel-2 wake dataset.

In addition to YOLOv26s, three other YOLO-based models were tested as benchmarks: YOLOv11s initialized with COCO-pretrained weights, YOLOv11s initialized with Sentinel-2 pretrained weights, and YOLOv8s. The following subsections describe the training setup and evaluation metrics using YOLOv26s as the main example. The parameter settings of the other benchmark models are provided in Appendix B.

4.4.2 Setup

During training, the model learned to predict the location of wake regions from satellite image tiles. The main training settings included the input image size, batch size, number of epochs, learning rate, optimizer, and data augmentation parameters. These settings are summarized in Table 4.5.

Setting	Value
Model	YOLOv26s
Pretraining	COCO-pretrained weights
Task	Single-class wake detection
Annotation format	Horizontal bounding box
Number of samples	Approximately 2000
Dataset split	70% - Train / 30% - validation
Image size	256*256 Pixels
Epochs	400
Batch size	16
Initial learning rate	0.001
Final learning rate	0.01
Optimizer	Auto
Augmentation	flipping / rotation / scaling / translation

Table 4.5: Training settings for the YOLO-based wake detection model.

During inference, the trained YOLO model was applied to Sentinel-2 image tiles to detect possible wake regions. The model output consists of bounding boxes, confidence scores, and class labels. For large satellite images, sliced inference can be used to reduce the loss of small wake details caused by image downscaling. The detected wake regions were then cropped from the original images and used in the following stages, including AIS matching and speed-wake dataset construction.

Several YOLO-based models and detection settings were compared under the same dataset split and evaluation protocol. This benchmark was designed to examine how different model choices affect wake detection performance in complex coastal satellite imagery.

4.5 Vessel Speed Prediction Framework

To estimate vessel speed from wake patterns observed in satellite imagery, a convolutional neural network regression framework was developed. The model was designed to learn the relationship between wake morphology and vessel speed using cropped Sentinel-2 wake images paired with AIS-derived speed labels. Unlike traditional analytical wake models requiring handcrafted hydrodynamic measurements, the proposed approach performs end-to-end feature extraction directly from image data.

4.5.1 Model

The proposed network consists of four convolutional stages followed by fully connected regression layers. The architecture was intentionally designed to remain computationally lightweight while retaining sufficient representational capacity for wake analysis. The complete architecture is summarized in Table 4.6.

Stage	Operation	Channels	Output Size
Input	RGB image	3	100 × 100
Block 1	Conv+BN+ReLU ×2 + MaxPool	32	50 × 50
Block 2	Conv+BN+ReLU ×2 + MaxPool	64	25 × 25
Block 3	Conv+BN+ReLU ×2 + MaxPool	128	12 × 12
Block 4	Conv+BN+ReLU + Adaptive Pool	256	4 × 4
FC Layers	Fully connected layers	128	–
Meta Branch	MLP on (w, h)	16	–
Output	Regression layer	1	Speed (kn)

Table 4.6: CNN architecture for vessel speed regression

The prediction pipeline consists of four major stages: dataset preparation, stratified data splitting, CNN-based feature learning, and regression-based speed estimation. The wake crops were generated from Sentinel-2 imagery using YOLO detections followed by AIS association. To preserve wake structures while maintaining consistent input dimensions, all crops were zero-padded rather than geometrically stretched. This prevents distortion of wake geometry, which is critical for preserving physically meaningful wake characteristics.

Each input sample consists of a fixed-size RGB wake image crop of size 100×100 pixels, auxiliary geometric metadata, normalized bounding-box width w, and normalized bounding-box height h. Pixel values x of the images were normalized to the range:

$$x_{norm} = \frac{x}{127.5} - 1 \tag{4.1}$$

which results in input values within $[-1, 1]$.

The feature extractor contains stacked convolutional blocks Conv, BatchNorm, and ReLU, followed by max-pooling operations. The network progressively increases channel depth. Adaptive average pooling was used to produce fixed-dimensional feature vectors independent of small spatial variations.

In addition to image features, normalized bounding-box dimensions were used as auxiliary inputs. Bounding-box size provides weak geometric priors related to vessel scale and wake extent. A small multilayer perceptron processes metadata:

$$[w, h] \to \mathbb{R}^{16} \tag{4.2}$$

where w is the normalized bounding-box wid is the th, and h is the normalized bounding-box height. The resulting vector is concatenated with image features before final regression. This multimodal fusion allows the model to combine visual wake structures with geometric context.

The final regression module predicts continuous vessel speed values in knots:

$$\hat{v} = g(z) \tag{4.3}$$

where z denotes the combined feature representation. The regression head consists of fully connected layers, ReLU activations, and dropout regularization. Dropout was introduced to reduce overfitting caused by the relatively limited dataset size.

4.5.2 Setup

The model was trained using the AdamW optimiser with initial learning rate $\eta = 10^{-3}$ and weight decay $\lambda = 10^{-4}$. The learning rate was annealed using cosine scheduling over a maximum of 150 epochs. Gradient norms were clipped to 5.0 to stabilise training under the weighted sampler, which stabilized the optimization. Training was halted by early stopping when the validation loss failed to improve by more than 10^{-3} for 20 consecutive epochs. The model checkpoint with the lowest validation loss was restored for test evaluation. The dataset was split into train/validation/test subsets at 70/15/15 using stratified sampling to preserve the speed distribution across all three splits. All experiments were conducted on a single NVIDIA GeForce RTX 4060. The training configuration is summarized in Table 4.7.

Parameter	Value
Input size	$100 \times 100 \times 3$
Batch size	32
Optimizer	AdamW
Learning rate	10^{-3}
Weight decay	10^{-4}
Loss function	Huber Loss ($\delta = 2.0$)
Scheduler	Cosine Annealing
Maximum epochs	150
Early stopping patience	20
Dropout	0.4
Gradient clipping	5.0

Table 4.7: CNN training configuration

4.6 Underwater Noise Map

The objective of this stage was to generate a spatially resolved indicator map showing the relative distribution of vessel-generated underwater noise within the study area.The vessel speed prediction method is further applied to a simplified underwater noise spatial indicator model for this purpose. This part of work is used as a simplified case study instead of producing a fully calibrated acoustic simulation.

The workflow combines two sources of vessel speed information: AIS-matched vessels with directly available speed measurements, and YOLO-detected vessels without AIS coverage, whose speeds are estimated by the proposed CNN wake-regression model. For each vessel, a simplified source level model was applied to convert vessel speed into an estimated acoustic emission level.

The parameters used in the implementation are shown in 4.8. This parameterization approximates the commonly used $N \propto V^4$ relationship observed in small recreational motorboats and planing vessels.

The study area was discretized into a $600*600$ spatial grid in geographic coordinates. Vessel positions were extracted from the AIS matching pipeline and converted into approximate metric distances using an equirectangular projection approximation. The resulting SPL field was visualized on top of high-resolution satellite imagery using Cartopy and ESRI World Imagery tiles. AIS-verified vessels were represented using green triangular markers, while CNN-estimated vessels were displayed as orange diamond markers. A semi-transparent heatmap overlay was used to represent cumulative SPL intensity across the study region.

The generated map therefore combines satellite vessel detection, AIS-supported speed estimation, CNN-based wake regression, simplified underwater noise propagation, and spatial visualization of cumulative vessel noise.

Parameter	Value
Reference source level SL_{ref}	140 dB re 1 μPa @ 1m
Reference speed V_{ref}	10 kn
Speed scaling coefficient α	60
Absorption coefficient β	0.001 dB/m
Grid resolution	600×600
SPL visualization range	70–130 dB
Basemap	ESRI World Imagery

Table 4.8: Noise mapping model parameters

5

Results and Discussion

This chapter presents the results obtained from the proposed framework. It first summarizes the constructed datasets, followed by the results of wake detection, speed estimation, and simplified noise mapping. The results are then discussed to evaluate the performance, limitations, and practical relevance of the proposed method.

5.1 Wake Detection

5.1.1 Performance

This section presents the YOLOv26s detection results on full-scene Sentinel-2 test images using SAHI. Since the model was trained on 256×256 patches, sliced inference was applied to larger test images to maintain detection performance for small wake regions. The evaluation includes quantitative results based on manually annotated ground-truth boxes and qualitative examples under different scene conditions.

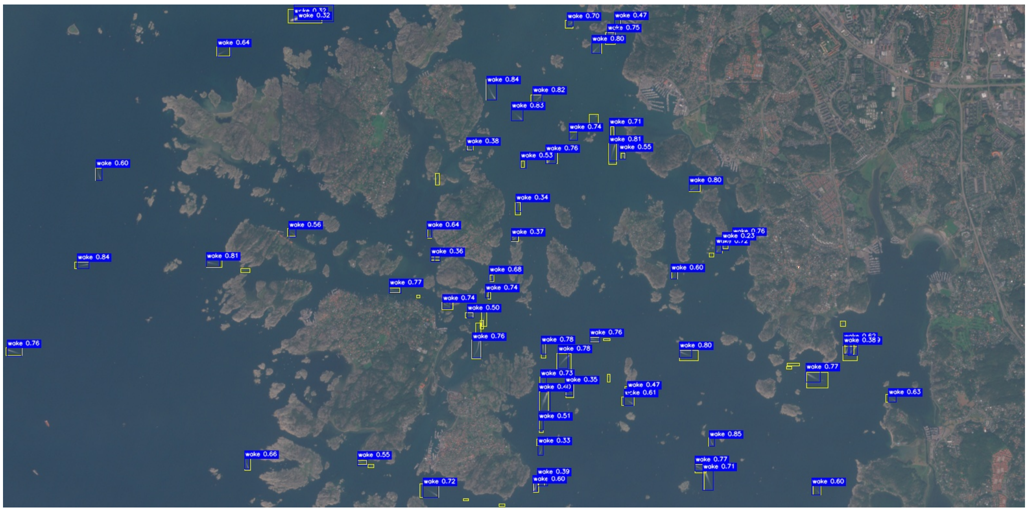


Figure 5.1: Comparison between YOLOv26s with SAHI predictions and manual ground-truth wake annotations on a representative test scene. Yellow boxes represent ground-truth wakes, while blue boxes represent predicted wake regions.

The representative test scene shows that YOLOv26s with SAHI can detect most visible wake regions across the full Sentinel-2 image. The predicted bounding boxes show good spatial agreement with the manually annotated ground-truth boxes. In many cases, the predicted boxes closely enclose the wake regions, with limited empty space between the wake structure and the box boundary. The confidence scores are also generally high, suggesting that the model can identify clear wake patterns with relatively strong certainty. The corresponding quantitative performance is reported in Table 5.1.

Image	GT wakes	Pred. wakes	TP	FP	FN	Precision	Recall	F1-score
Fig. 5.1	79	64	60	4	19	0.938	0.759	0.839

Table 5.1: Detection performance for the representative test scene.

Table 5.2 summarizes the overall detection performance of YOLOv26s with SAHI on the test set. The model achieved a precision of 0.859, recall of 0.759, and F1-score of 0.806. These results show that the detector performs reliably on full-scene Sentinel-2 images and can identify most visible wake regions after sliced inference. The relatively high precision indicates that most predicted wake boxes correspond to true wake structures, while the lower recall suggests that some wakes are still missed. Overall, the results indicate that YOLOv26s with SAHI provides a practical detection performance for satellite-based wake extraction, although further improvement is still needed for faint and complex wake cases.

TP	FP	FN	Precision	Recall	F1-score	IoU	Conf.	Slice size	Overlap
390	64	124	0.859	0.759	0.806	0.2	0.2	256 × 256	0.2, 0.2

Table 5.2: Overall performance of YOLOv26s with SAHI on the test set.

Figure 5.2: Wake detection in a crowded waterway with complex background interference, including sea-surface clutter, partial cloud cover, complex coastlines, dense vessel traffic, and overlapping wakes.

In the crowded waterway scene, detection performance decreases compared with the representative example. The precision, recall, and F1-score are 0.779, 0.744, and 0.761, respectively. This suggests that dense traffic and complex background patterns increase both false positives and missed detections.

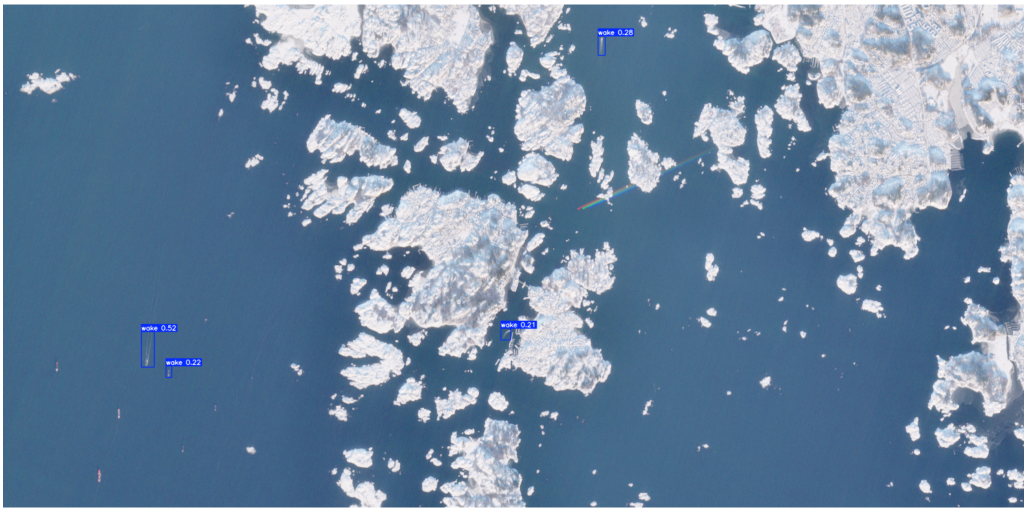


Figure 5.3: Wake detection under snowy background conditions, where bright snow-covered land areas, an aircraft trajectory, and vessels without clear wakes introduce additional visual distractions.

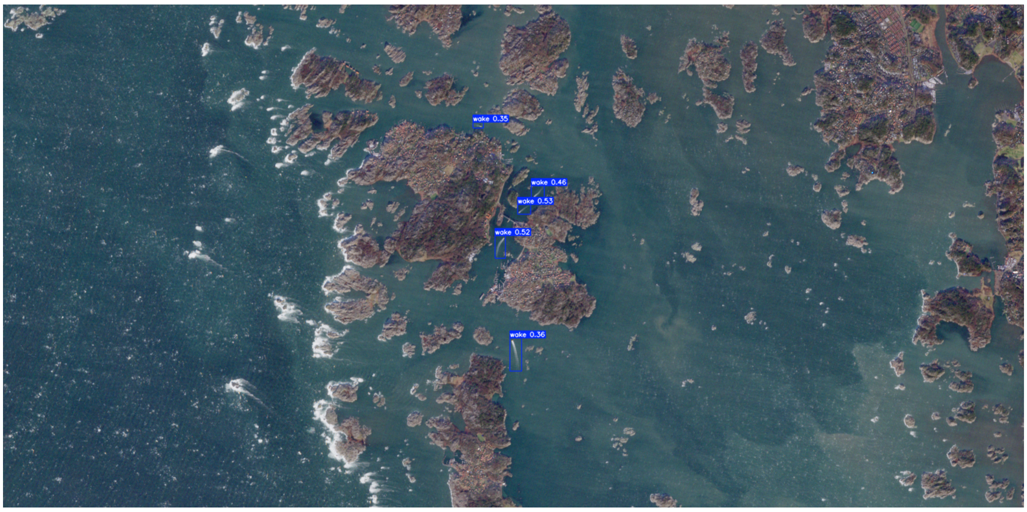


Figure 5.4: Wake detection with strong sea-surface clutter, breaking waves, coastlines, and visible background texture.

The results in Figures 5.3 and 5.4 show that the detector can still identify wake regions under challenging environmental conditions. However, wake-like background patterns may increase the risk of false detections and missed detections.

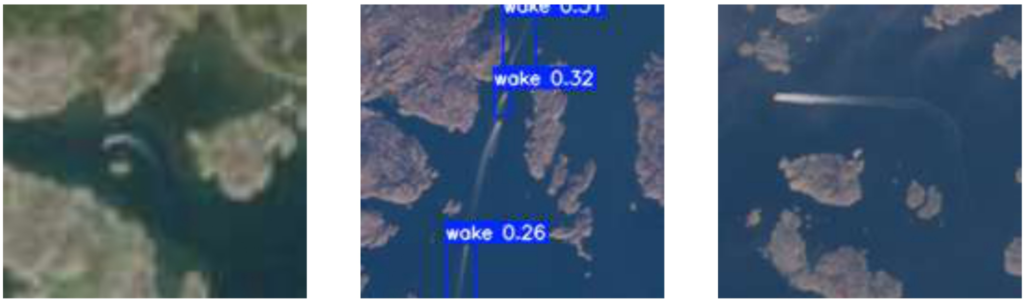
Image	GT wakes	Pred. wakes	TP	FP	FN	Precision	Recall	F1-score
Fig. 5.2	90	86	67	19	23	0.779	0.744	0.761
Fig. 5.3	4	4	4	0	0	1.0	1.0	1.0
Fig. 5.4	5	5	5	0	0	1.0	1.0	1.0

Table 5.3: Detection performance under challenging scene conditions.

5.1.2 Limitations

Although the model shows good performance in several test scenes, some failure cases are still observed. False positives mainly occur in areas with wake-like background patterns. These regions can have similar brightness and linear textures to vessel wakes. False negatives are more likely when wakes are weak, short, low-contrast, or partially mixed with the background. In crowded waterways, overlapping wakes and dense vessel traffic can also make localization more difficult. In some cases, the predicted boxes only cover part of the wake region or show a small spatial offset from the ground-truth annotation.

cloud cover cloud cover cloud cover / sea clutter

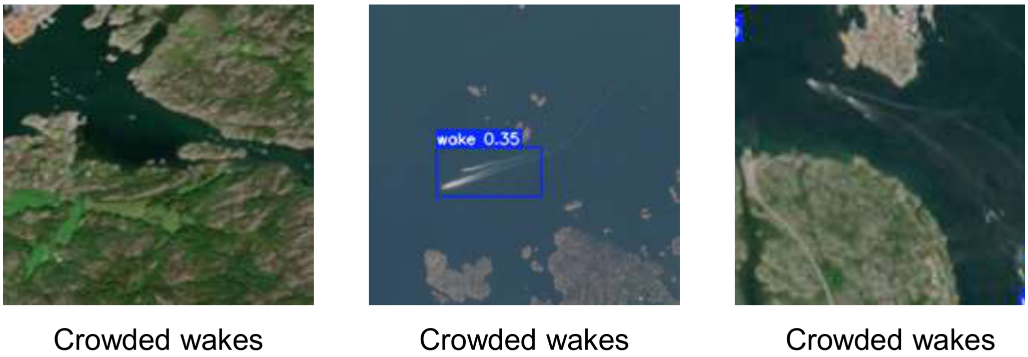


irregular wake irregular wake irregular wake

Figure 5.5: Examples of YOLOv26s wake detection failure cases. The first row shows false negatives under cloud cover and sea-surface clutter, where wake visibility and contrast are reduced. The second row shows irregular wake patterns, including curved, turning, or angular wake structures that differ from common straight wake patterns.

As shown in Figure 5.5, cloud cover and sea-surface clutter can reduce the visibility of wake structures and increase the probability of missed detections. Based on the test observations, detection accuracy is noticeably affected when cloud cover becomes substantial, especially above approximately 30%. Wakes partially covered by clouds or mixed with bright cloud edges are more likely to be missed.

Irregular wake patterns also remain difficult for the model. Since curved, turning, or angular wakes are relatively rare in the annotated data, the model has fewer examples from which to learn these visual patterns. As a result, unusual wake structures are not detected as reliably as straight and elongated wakes.



Crowded wakes Crowded wakes Crowded wakes

Figure 5.6: Examples of YOLOv26s wake detection failure cases in crowded waterways. Multiple nearby vessels and overlapping wake traces can form complex visual patterns, causing missed detections, merged predictions, or bounding boxes that only cover part of the dominant wake region.

Another common limitation is related to crowded or overlapping wakes, as shown in Figure 5.6. In dense waterways, several vessels may travel close to each other, and their wakes can overlap or merge visually in Sentinel-2 imagery. This can cause the model to miss individual wake regions, detect only the most visible wake, or merge several nearby wakes into one prediction. These results indicate that crowded wake patterns remain challenging, especially when wake boundaries are weak or visually mixed with nearby wake traces.

5.2 Vessel Speed Prediction

5.2.1 Performance

The proposed CNN-based wake regression framework demonstrated stable optimization behavior throughout training. The convergence of the training curves in 5.7 indicates successful feature learning from wake morphology and surrounding water textures.

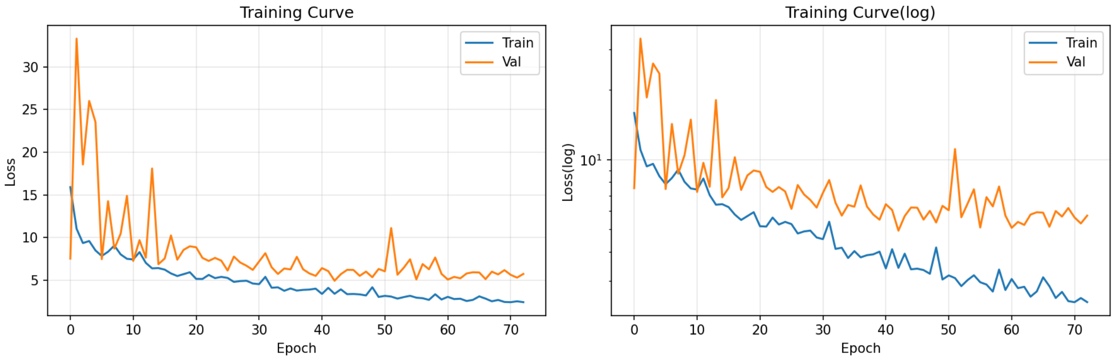


Figure 5.7: Training curves of CNN. Both training and validation losses decreased rapidly during the early epochs. Early stopping terminated the training process at epoch 73 to mitigate overfitting and preserve the model state with the best validation performance.

The final model achieved a MAE of 2.74 kn and a RMSE of 4.34 kn on the held-out test set (n=96). Compared with a statistical mean baseline predictor using the mean vessel speed of the training set, the proposed method reduced prediction error by approximately 62.6%. These results indicate that wake morphology extracted from Sentinel-2 imagery contains meaningful information related to vessel dynamics and operational speed.

Table 5.4: Overall regression performance on the test set

Method	MAE (kn)	RMSE (kn)	Relative Error vs CNN
Mean baseline predictor	7.33	–	+167%
Proposed CNN	2.74	4.34	–

The prediction-versus-ground-truth scatter plot shown in further demonstrates that the model successfully captured the overall nonlinear relationship between wake appearance and vessel speed.

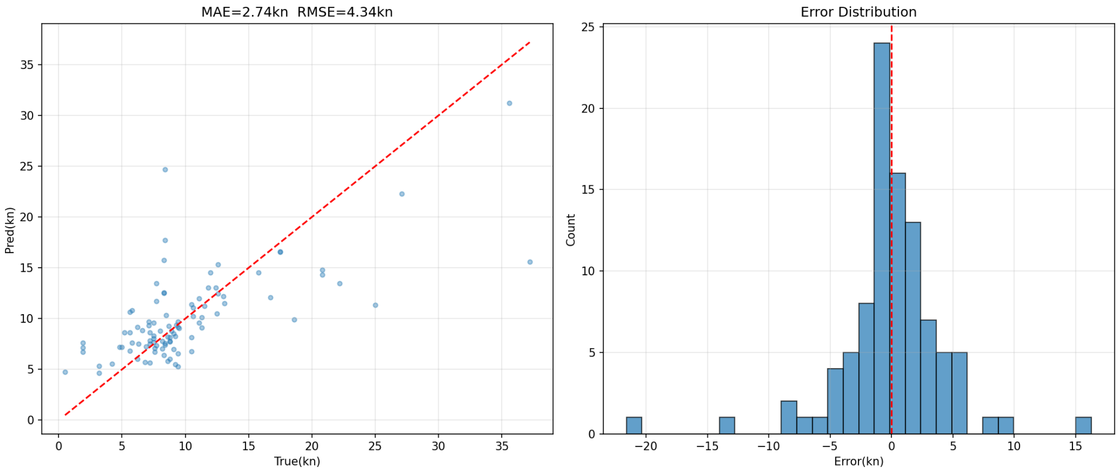


Figure 5.8: Prediction scatter plot and error distribution of speed estimation. Most predictions clustered near the identity line, particularly within the dominant interval of 7-12 kn, which corresponds to the most common speed range in the dataset. Larger prediction deviations were observed at the extreme ends of the speed distribution.

To better evaluate performance under different operational conditions, the test set was additionally analyzed using speed-stratified evaluation. The CNN achieved its best performance within the $7 - 12$ kn interval, reaching an MAE of only 2.02 kn. This region also contained the largest number of samples, suggesting that the model benefited significantly from the availability of sufficient training examples and more consistent wake patterns.

Prediction accuracy degraded substantially for vessels above 25 kn, where the MAE increased to 11.10 kn. However, only four samples were available in this interval, making the evaluation highly sensitive to individual prediction errors. High-speed wakes additionally exhibit greater variability due to planing behavior, strong turbulence, spray generation, and varying wake geometries, all of which complicate regression from single optical images.

Table 5.5: Speed-stratified regression performance on test set

Speed Range (kn)	Number of Samples	MAE (kn)
0–3	4	5.02
3–7	17	2.19
7–12	56	2.02
12–25	15	3.21
25–40	4	11.10

The visualization in Figure 5.9 further demonstrates the practical applicability of the proposed framework. The model enables the reconstruction of spatial vessel activity directly from satellite imagery. This capability is particularly relevant because many small recreational vessels are not legally required to carry AIS transponders, resulting in incomplete coverage in traditional maritime monitoring systems.

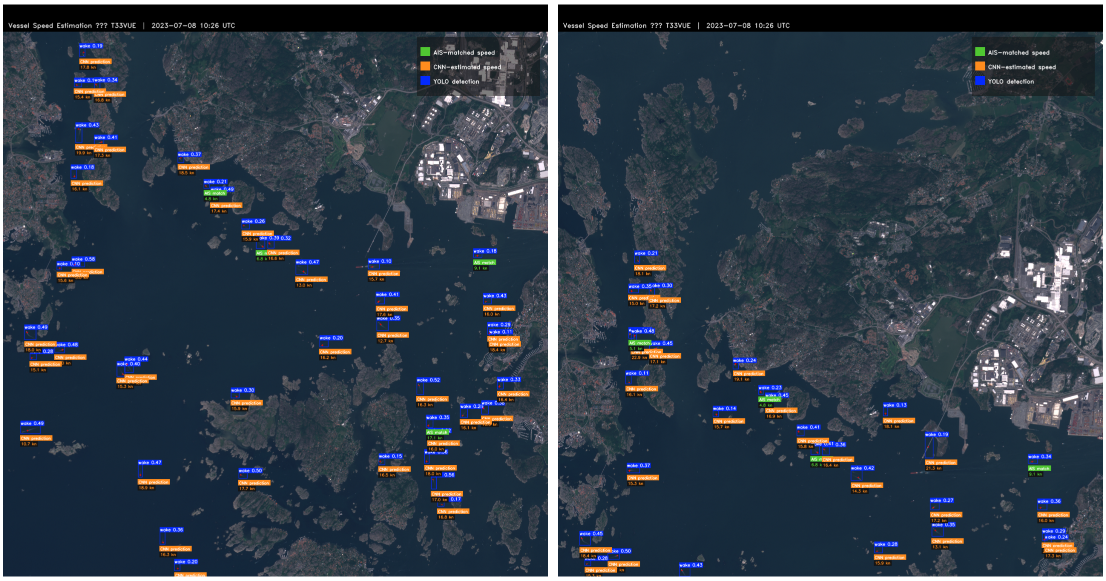


Figure 5.9: YOLO detection and speed estimation results around the Kattegat Sea area (Sentinel-2 Tile T33VUE) on 2023-07-08. The model successfully estimated plausible vessel speeds for numerous recreational boats without AIS coverage.

5.2.2 Comparisons

In addition to the statistical baseline comparison, several traditional machine learning regressors were implemented to evaluate whether deep spatial wake features provide advantages beyond simple geometric descriptors. Feature embeddings extracted from the trained CNN backbone were used as inputs to linear regression, random forest regression, and gradient boosting regression models.

The CNN consistently outperformed all baseline methods, indicating that convolutional spatial feature extraction provides superior representation of complex wake morphology compared with conventional regression approaches.

Table 5.6: Comparison between the proposed CNN and baseline regression models

Method	MAE (kn)	Relative Error vs CNN
Mean baseline predictor	7.33	+167%
Linear Regression	4.53	+65%
Random Forest Regression	3.63	+32%
Gradient Boosting Regression	3.82	+39%
Proposed CNN	2.74	–

The results suggest that the learned convolutional representations capture meaningful wake characteristics associated with vessel motion. Unlike traditional regressors relying primarily on global statistics or low-dimensional descriptors, the CNN can jointly model wake texture, Kelvin wake structure, turbulent intensity, spatial brightness distribution, and contextual water patterns. This ability likely explains the improved regression performance observed across the majority of operational speed ranges.

5.2.3 Limitations

Despite the encouraging results, several important limitations must be considered. First, Sentinel-2 imagery provides relatively limited spatial resolution (10m per pixel) and revisit frequency compared with dedicated airborne or SAR systems. Small wake structures may therefore occupy only a limited number of pixels, restricting the amount of recoverable information.

Second, AIS-derived labels contain unavoidable temporal synchronization uncertainty. Vessel positions recorded in AIS streams are not perfectly aligned with the satellite acquisition timestamp, particularly for fast-moving recreational boats. Small temporal offsets may introduce non-negligible position and speed inconsistencies during the matching process.

Third, wake appearance is influenced by many additional physical variables beyond vessel speed alone. These include vessel hull geometry, displacement and draft, propulsion type, sea state and wind conditions, viewing geometry, sun glint, and local water surface roughness.

Consequently, the mapping between wake appearance and vessel speed is inherently noisy and partially ambiguous when inferred from single optical images.

The long-tailed speed distribution of the dataset further constrained performance at extreme speed intervals. Medium-speed recreational boats dominated the dataset, whereas very slow and very fast vessels were comparatively rare. Although weighted sampling and data augmentation partially alleviated this imbalance, the limited number of high-speed training examples remained a major bottleneck.

Nevertheless, the proposed framework substantially outperformed both statistical and conventional machine learning baselines while demonstrating the feasibility of vessel-speed estimation directly from satellite wake imagery. The results therefore support the broader hypothesis that optical wake signatures contain exploitable physical information related to vessel dynamics and environmental impact.

From an application perspective, the framework additionally enables downstream environmental analyses such as spatial noise mapping and maritime activity monitoring in regions where AIS coverage is incomplete or unavailable.

5.3 Underwater Noise Map

A case study was conducted using Sentinel-2 imagery acquired around the Kattegat Sea area on 8 July 2023. The selected scene contains a number of recreational vessels operating within narrow coastal channels and island passages.

The vessel-speed estimation framework first generated vessel detections using YOLOv26. Vessels carrying AIS transponders were matched directly using spatio-temporal constraints, while remaining detections were processed using the CNN wake-regression model. The resulting speed predictions were subsequently converted into acoustic source levels and propagated spatially using the proposed indicator model.

Figure 5.10 demonstrates the noise map of the study area. Several vessels without AIS coverage were successfully assigned plausible speed estimates solely from wake appearance, illustrating the ability of the CNN model to supplement incomplete AIS observations.

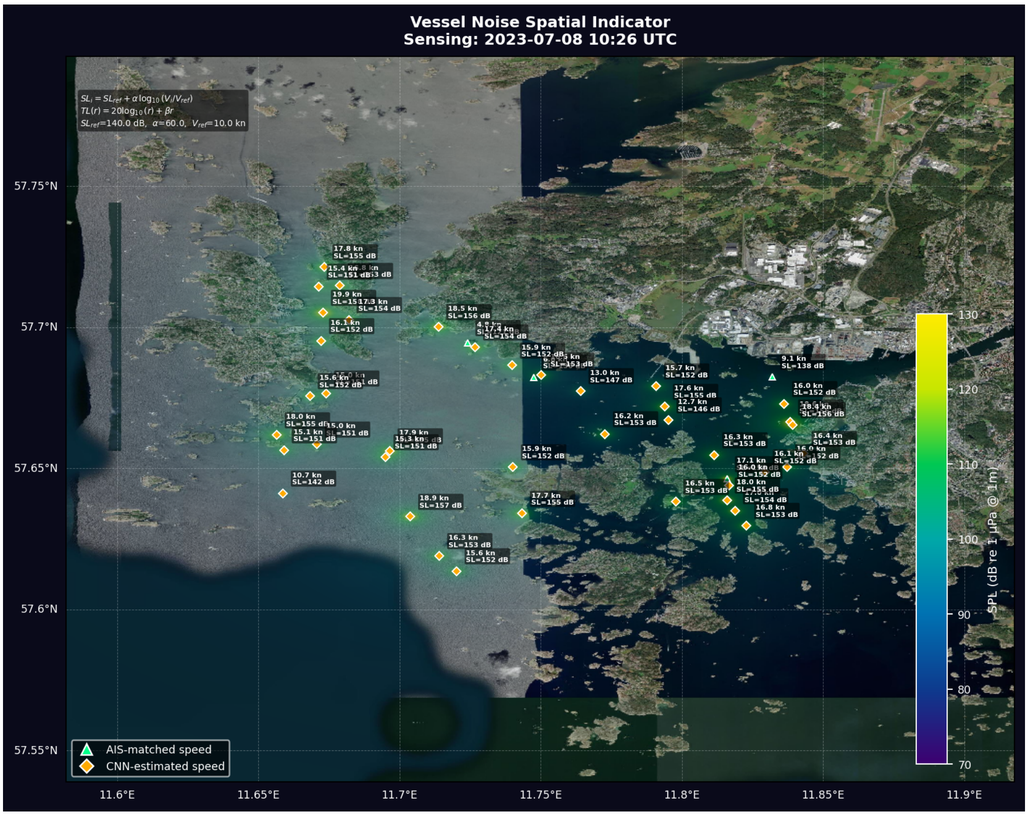


Figure 5.10: Noise map of a Sentinel-2 image around the Kattegat Sea area (Sentinel-2 Tile T33VUE) on 2023-07-08. The majority of detected vessels were estimated within the 10 – 20 kn operational range, which is consistent with typical recreational boat activity in west Swedish archipelagos.

The noise distribution of the result is shown in Figure 5.11. The resulting spatial indicator map reveals several localized acoustic hotspots concentrated along high-traffic navigation corridors. Regions with dense vessel clustering exhibit overlapping SPL fields caused by incoherent energy accumulation from multiple nearby vessels.

The generated map also demonstrates a key practical contribution of the proposed framework: recreational boats without AIS coverage can still contribute to spatial noise estimation through satellite-based wake analysis. This capability is particularly relevant because many small recreational vessels operating in coastal waters are not legally required to carry AIS transmitters.

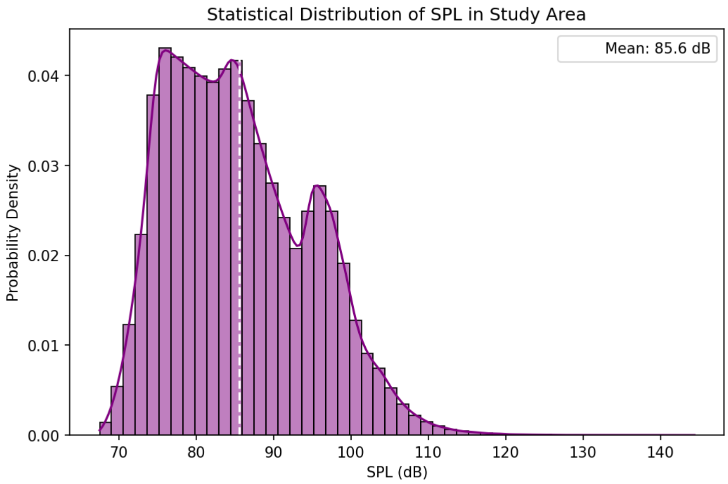


Figure 5.11: SPL distribution of the study area. The strongest acoustic contributions were primarily associated with medium-to-high-speed recreational motorboats whose estimated source levels exceeded 150 dB re 1 μPa @ 1m.

Although the proposed model does not attempt full acoustic calibration, the results demonstrate the feasibility of integrating satellite remote sensing, vessel detection, wake-based speed regression, and simplified propagation modeling into a unified environmental monitoring framework.

Therefore, the obtained SPL field should be interpreted as a spatial indicator of relative vessel-noise intensity rather than an absolute underwater acoustic prediction. Several factors constrain physical accuracy, including simplified free-field propagation assumptions, absence of bathymetry and seabed modeling, unknown vessel engine characteristics, single-image temporal sampling, and uncertainty in AIS-derived labels.

Nevertheless, the produced ttindicator map successfully highlights the spatial structure of vessel activity and demonstrates the applicability of CNN-based wake-speed estimation to environmental marine monitoring tasks.

6

Conclusion

This thesis investigated the use of machine learning based methods for recreational vessel wake detection from satellite imagery, speed estimation, and preliminary speed-related noise mapping. The end-to-end workflow provides an AI-based solution in the maritime domain. Sentinel-2 optical imagery was selected as the main data source due to its suitable balance between spatial resolution, spatial coverage, revisit frequency, and open accessibility. Based on Sentinel-2 images, a manually annotated wake dataset was constructed for training and evaluating YOLO-based object detection models.

The final detection workflow used YOLOv26s combined with SAHI to detect small wake regions in full-scene Sentinel-2 images. The model was tested under different coastal and environmental conditions, including crowded waterways, snow-covered backgrounds, cloud interference, and strong sea-surface clutter. The overall test performance reached a precision of 0.859, recall of 0.759, and F1-score of 0.806. These results show that the proposed workflow can localize most visible wake regions, although weak, irregular, and overlapping wakes remain challenging.

The detected wake regions were further matched with AIS vessel records to generate a wake–speed dataset. A CNN-based regression model was then trained to estimate vessel speed from wake image crops. The model achieved a test-set MAE of 2.74 kn and outperformed statistical and traditional machine learning baselines, including linear regression, random forest regression, and gradient boosting regression. This indicates that Sentinel-2 wake morphology contains useful visual information related to vessel speed, even under limited spatial resolution.

Finally, a simplified vessel-noise mapping framework was developed using the estimated vessel speeds. Although the acoustic model is simplified and mainly serves as a proof of concept, the results demonstrate the potential of combining satellite imagery, computer vision, AIS data, CNN-based regression, and environmental modelling for large-scale coastal vessel monitoring and preliminary underwater noise assessment.

Overall, this thesis demonstrates the feasibility of using open-access Sentinel-2 imagery and machine learning methods for recreational vessel wake detection, speed estimation, and speed-related environmental monitoring in coastal waters, and shows the potential of AI application in the maritime domain.

6.1 Future Work

Several directions can be explored to improve the proposed workflow. First, the wake detection dataset should be expanded with more diverse samples from different seasons, sea states, lighting conditions, weather conditions, and geographic regions. More examples of cloud cover, sea-surface clutter, snow-covered backgrounds, weak wakes, curved wakes, and overlapping wakes would help improve the robustness of the detector. Since overlapping wakes share similar challenges with crowded object detection, future work could also investigate data augmentation methods for dense or occluded patterns, such as copy-paste-based augmentation [37, 38]. The model should also be tested on other image sources, such as Landsat, commercial high-resolution satellite imagery, or drone imagery, to evaluate whether it can generalize beyond Sentinel-2 true-colour images.

Second, the AIS–satellite matching and speed estimation steps can be improved. The current matching method may introduce temporal and spatial uncertainty because the satellite sensing time may not exactly correspond to each vessel position in the image. Future work could use AIS trajectory interpolation, vessel heading, and speed information to obtain more accurate wake–vessel pairs. For speed regression, a larger and more balanced wake–speed dataset is needed, especially for very slow and very fast vessels. Additional information, such as vessel size, heading, wake orientation, sea state, wind conditions, and multi-temporal trajectories, could also be incorporated to improve the accuracy and robustness of the regression model.

Third, the noise mapping framework can be further developed with more realistic acoustic modelling. The current framework is mainly a proof of concept and uses simplified assumptions for source level estimation and sound propagation. Future work could include vessel type, propulsion system, bathymetry, coastline geometry, water temperature, salinity, background noise, and frequency-dependent transmission loss. Repeated satellite observations could also be used to analyse long-term spatial noise exposure patterns in coastal ecosystems. In the longer term, satellite-based wake detection and speed estimation may provide complementary information for maritime monitoring in areas where AIS coverage is incomplete, especially for recreational vessels.

Bibliography

[1] E. M. S. Agency, “Navison final report: Calculation and analysis of shipping sound maps for all european seas from 2016 to 2050,” EMSA, Lisbon, Tech. Rep., 2024. [Online]. Available: https://www.emsa.europa.eu/navison.html

[2] J. Mäyrä *et al.*, “Mapping recreational marine traffic from Sentinel-2 imagery using YOLO object detection models,” *Remote Sensing of Environment*, vol. 326, p. 114791, 2025.

[3] D. P. Nowacek, L. H. Thorne, D. W. Johnston, and P. L. Tyack, “Responses of cetaceans to anthropogenic noise,” *Mammal Review*, vol. 37, no. 2, pp. 81–115, 2007.

[4] M. Zucchetta, F. Madricardo, M. Ghezzo, A. Petrizzo, and M. Picciulin, “Satellite-based monitoring of small boat for environmental studies: A systematic review,” *Journal of Marine Science and Engineering*, vol. 13, no. 3, p. 390, 2025.

[5] G. Zilman, A. Zapolski, and M. Marom, “The speed and beam of a ship from its wake’s SAR images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2335–2343, 2004.

[6] L. Wang, B. Sun, C. Zhao, T. Ohtsuki, and P. T. Mathiopoulos, “SAR ship wake detection based on siamese network with Mamba cross-domain feature fusion,” in *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2026, pp. 11 392–11 396.

[7] T. Torsvik, “Modelling of ship waves from high-speed vessels,” in *Applied Wave Mathematics*, E. Quak and T. Soomere, Eds., 2009, pp. 229–263.

[8] M. Nastasi *et al.*, “Parameters affecting noise emitted by ships moving in port areas,” *Sustainability*, vol. 12, no. 20, p. 8742, 2020.

[9] C. A. F. de Jong, B. Binnerts, and M. A. Ainslie, “Jomopans north sea sound maps 2019–2020,” Interreg North Sea Region Programme, Tech. Rep., 2022.

[10] A. Mazzeo, A. Renga, and M. D. Graziano, “A systematic review of ship wake detection methods in satellite imagery,” *Remote Sensing*, vol. 16, no. 20, p. 3775, 2024.

[11] K.-M. Kang and D.-J. Kim, “Ship velocity estimation from ship wakes detected using convolutional neural networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 11, pp. 4379–4388, 2019.

[12] Y. Liu and R. Deng, “Ship wakes in optical images,” *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 8, pp. 1633–1648, 2018.

[13] R. Magalhães *et al.*, “Vessel detection leveraging satellite imagery and YOLO in maritime surveillance,” *Remote Sensing Applications: Society and Environment*, vol. 40, p. 101730, 2025.

[14] R. Qiu, N. Bi, and C. Yin, “Optwake-YOLO: a lightweight and efficient ship wake detection model based on optical remote sensing images,” *Frontiers in Marine Science*, vol. 12, p. 1624323, 2025.

[15] M. Gao *et al.*, “Ship-VNet: An algorithm for ship velocity analysis based on optical remote sensing imagery containing Kelvin wakes,” *Electronics*, vol. 13, no. 17, p. 3468, 2024.

[16] K. Haakman *et al.*, “Sea surface current estimation using optical satellite imagery of Kelvin wakes and AIS data,” *Remote Sensing of Environment*, vol. 315, p. 114426, 2024.

[17] L. Hermannsen *et al.*, “Recreational vessels without Automatic Identification System (AIS) dominate anthropogenic noise contributions to a shallow water soundscape,” *Scientific Reports*, vol. 9, no. 1, p. 15477, 2019.

[18] Ultralytics, “Ultralytics YOLO26,” 2026. [Online]. Available: https://docs.ultralytics.com/models/yolo26/

[19] P. Hidayatullah, “YOLO26: A comprehensive architecture overview and key improvements,” 2026, arXiv preprint.

[20] R. Sapkota and M. Karkee, “Ultralytics YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 object detectors,” 2025, arXiv:2510.09653.

[21] Ultralytics, “Understanding end-to-end detection in ultralytics YOLO26,” 2026. [Online]. Available: https://docs.ultralytics.com/guides/end2end-detection/

[22] F. C. Akyon, S. O. Altinuc, and A. Temizel, “Slicing aided hyper inference and fine-tuning for small object detection,” in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 966–970.

[23] OpenCV, “Morphological transformations,” 2024. [Online]. Available: https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html

[24] scikit-image, “Removing small objects in grayscale images with a top hat filter,” 2024. [Online]. Available: https://scikit-image.org/docs/stable/auto_examples/filters/plot_tophat.html

[25] A. M. Reed and J. H. Milgram, “Ship wakes and their radar images,” *Annual Review of Fluid Mechanics*, vol. 34, pp. 469–502, 2002.

[26] H. Xiong *et al.*, “Passive millimeter-wave polarization imaging of ship Kelvin wake on the sea surface,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 64, pp. 1–16, 2026.

[27] Y. LeCun and Y. Bengio, “Convolutional networks for images, speech, and time series,” in *The Handbook of Brain Theory and Neural Networks*, 1998, hAL Id: hal-05083427.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.

[29] J. T. Barron, “A general and adaptive robust loss function,” 2019, arXiv:1701.03077.

[30] A. Affatati, C. Scaini, and S. Salon, “Ocean sound propagation in a changing climate,” *Earth’s Future*, vol. 10, no. 3, p. e2021EF002099, 2022.

[31] C. D. S. Ecosystem, “Sentinel-1,” 2024. [Online]. Available: https://dataspace.copernicus.eu/data-collections/copernicus-sentinel-missions/sentinel-1

[32] E. S. Agency, “Sentinel-2,” 2024. [Online]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2

[33] ——, “Sentinel-3 instruments,” 2024. [Online]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-3/Instruments

[34] U. G. Survey, “Landsat 8,” 2024. [Online]. Available: https://www.usgs.gov/landsat-missions/landsat-8

[35] K. Nordberg *et al.*, “Tracing PAH emissions from leisure boats in a low tidal environment,” *Chemosphere*, 2025.

[36] I. P. Asteman *et al.*, “Leisure boat harbours, hidden alien species, and pollution,” *Marine Pollution Bulletin*, 2025.

[37] S. Shao *et al.*, “Crowdhuman: A benchmark for detecting human in a crowd,” 2018, arXiv:1805.00123.

[38] J. Deng, D. Fan, X. Qiu, and F. Zhou, “Improving crowded object detection via copy-paste,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, 2023, pp. 497–505.

A

Appendix 1

A.1 YOLOv26s with SAHI Inference Algorithm

Algorithm 1: YOLOv26s with SAHI sliced inference workflow

```
Input: Input image $I$, trained YOLOv26s model $M$, confidence threshold
       $c$, slice size $S$, overlap ratio $r$
Output: Merged wake detections $D$ and visualization image
Load trained YOLOv26s model $M$;
Set confidence threshold $c = 0.2$;
Set slice size $S = 256 \times 256$;
Set overlap ratio $r = 0.2$;
Divide the input image $I$ into overlapping image slices;
for each slice $s_i$ do
    Apply YOLOv26s model $M$ to detect wake candidates;
    Remove low-confidence predictions below threshold $c$;
    Store bounding boxes, class labels, and confidence scores;
Map all slice-level detections back to the coordinate system of the original
  image;
Apply post-processing to merge overlapping detections from adjacent slices;
Let the final merged detections be $D$;
for each detection $d_j \in D$ do
    Extract bounding box coordinates and confidence score;
    Draw the predicted bounding box on the original image;
    Add class label and confidence score for visualization;
Save final detection visualization;
return $D$;
```

A.2 Temporal-Spatial AIS Matching Algorithm

Algorithm 2: Temporal-Spatial AIS Matching

```
Input: Detected vessel bow positions $\mathcal{D}$ from satellite imagery;
AIS records $\mathcal{A}$ containing timestamp, latitude, longitude, MMSI and speed;
Satellite sensing time $t_s$;
Temporal threshold $\Delta t = 1$ minute;
Maximum matching distance $d_{\max} = 150$ m
Output: Matched AIS record for each detected vessel
Initialize empty matched set $\mathcal{M}$;
foreach detected vessel $d_i \in \mathcal{D}$ do
    Extract bow-point coordinate $(lat_i, lon_i)$;
    Select temporal AIS candidates:
        $$\mathcal{A}_t = \{a_j \in \mathcal{A} \mid |t_j - t_s| \leq \Delta t\}$$
    ;
    Filter AIS candidates inside the image geographic bounds:
        $$\mathcal{A}_s \subseteq \mathcal{A}_t$$
    ;
    foreach candidate $a_j \in \mathcal{A}_s$ do
        Compute Haversine distance:
            $$d_{ij} = 2R \arcsin \sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_i \cos\phi_j \sin^2\left(\frac{\Delta\lambda}{2}\right)}$$
        ;
    Find nearest candidate:
        $$a^* = \arg\min_{a_j \in \mathcal{A}_s} d_{ij}$$
    ;
    if $d(a^*) \leq d_{\max}$ then
        Assign $a^*$ to vessel $d_i$;
        Store match in $\mathcal{M}$;
foreach duplicate MMSI in $\mathcal{M}$ do
    Retain only the match with minimum distance;
Return final matched set $\mathcal{M}$;
```

A.3 Wake Head Localization Using Top-K Brightest Pixels

Algorithm 3: Wake-head localization using top-K brightest pixels.

```
Input: Original image $I$, YOLO detections $D$, number of brightest pixels $K$
Output: YOLO labels with confidence and estimated wake-head coordinates
Initialize three output label files: standard labels, confidence labels, and bright-$K$ labels;
for each detection $d_i \in D$ do
    Extract bounding box coordinates $(x_1, y_1, x_2, y_2)$;
    Extract class label $c_i$ and confidence score $s_i$;
    Convert bounding box coordinates to normalized YOLO format: $(x_c, y_c, w, h)$;
    Crop the image region inside the bounding box;
    Convert the cropped region to grayscale;
    Find the top-$K$ brightest pixels inside the cropped region;
    if bright pixels are found then
        Compute the brightness-weighted average position of the top-$K$ pixels;
        Convert this position back to the original image coordinate system;
        Normalize the estimated wake-head position by image width and height;
        Draw a marker at the estimated wake-head position for visual inspection;
    else
        Set the wake-head position to $(-1, -1)$;
    Save standard YOLO label: class, centre position, width, and height;
    Save confidence label: class, centre position, width, height, and confidence score;
    Save bright-$K$ label: class, centre position, width, height, confidence score, and estimated wake-head coordinates;
return standard labels, confidence labels, and bright-K labels;
```

A.4 Top-hat Feature Enhancement Algorithm

Algorithm 4: Top-hat feature enhancement workflow

```
Input: Input image crop $I$
Output: Enhanced image $I_{\text{enh}}$
Read input image $I$;
Convert $I$ from RGB to grayscale image $I_g$;
Apply bilateral filtering to reduce small-scale noise while preserving wake
  boundaries;
Let the denoised image be $I_b$;
Define a structuring element $B$ with selected kernel size;
Apply morphological opening to $I_b$ using $B$;
    $I_{\text{open}} = I_b \circ B$
Compute the white top-hat transform;
    $I_{\text{enh}} = I_b - I_{\text{open}}$
Normalize $I_{\text{enh}}$ to the range $[0, 255]$;
Save the enhanced image $I_{\text{enh}}$;
return $I_{\text{enh}}$;
```

B

Appendix 2

B.1 YOLOv26s Training Parameters

Table B.1: YOLOv26s training parameters

Parameter	Value
Model	YOLOv26s
Framework	Ultralytics 8.4.21
Python version	3.12.12
PyTorch version	2.10.0+cu128
GPU	NVIDIA GeForce RTX 4060
Task	Object detection
Epochs	400
Batch size	16
Input image size	960
Optimizer	auto
Initial learning rate	0.001
Final learning rate factor	0.01
Momentum	0.937
Weight decay	0.0005
IoU threshold	0.7
AMP	True
Pretrained weights	Official COCO-pretrained weights
Mosaic augmentation	0.0
Horizontal flip probability	0.5
Vertical flip probability	0.0
HSV augmentation	h=0.015, s=0.7, v=0.4
Translation	0.1
Scale	0.5
Workers	8
Seed	0

B.2 YOLOv11s Training Parameters

Table B.2: YOLOv11s training parameters

Parameter	Value
Model	YOLOv11s
Framework	Ultralytics 8.4.19
Python version	3.12.12
PyTorch version	2.10.0+cu128
GPU	NVIDIA GeForce RTX 4060
Task	Object detection
Epochs	180
Batch size	16
Input image size	960
Optimizer	auto
Initial learning rate	0.01
Final learning rate factor	0.01
Momentum	0.937
Weight decay	0.0005
IoU threshold	0.7
AMP	True
Pretrained weights	Custom pretrained weights
Mosaic augmentation	0.0
Horizontal flip probability	0.5
Vertical flip probability	0.2
HSV augmentation	h=0.0, s=0.0, v=0.0
Rotation degrees	10.0
Translation	0.1
Scale	0.3
Patience	30
Workers	8
Seed	0

B.3 YOLOv8s Training Parameters

Table B.3: YOLOv8s training parameters

Parameter	Value
Model	YOLOv8s
Framework	Ultralytics
Task	Object detection
Epochs	450
Batch size	16
Input image size	960
Optimizer	auto
Initial learning rate	0.001
Final learning rate factor	0.1
Momentum	0.937
Weight decay	0.0005
IoU threshold	0.7
AMP	True
Pretrained weights	Official COCO-pretrained weights
Mosaic augmentation	0.0
Horizontal flip probability	0.5
Vertical flip probability	0.2
HSV augmentation	h=0.0, s=0.0, v=0.0
Rotation degrees	10.0
Translation	0.1
Scale	0.3
Workers	8
Seed	0

C

Appendix 3

C.1 Training Curves of YOLO Benchmark Models

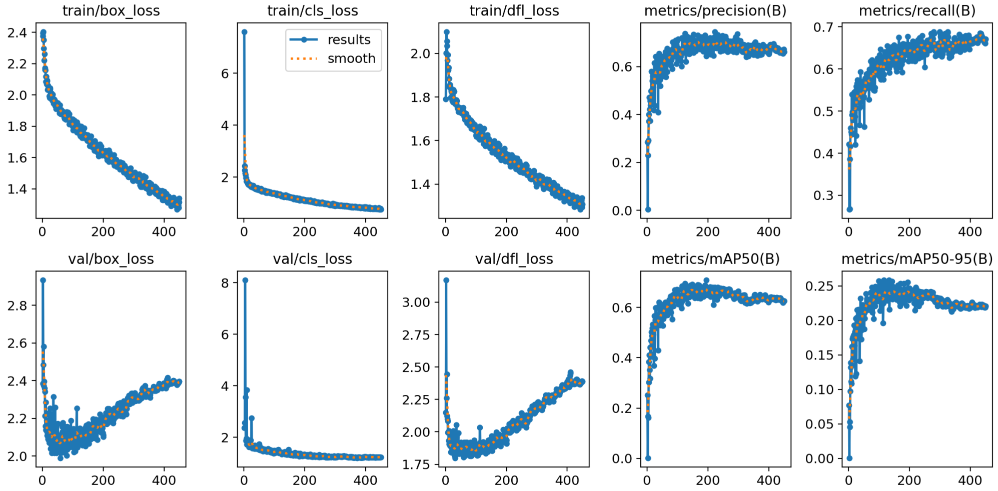


Figure C.1: YOLOv8s training curves with COCO dataset

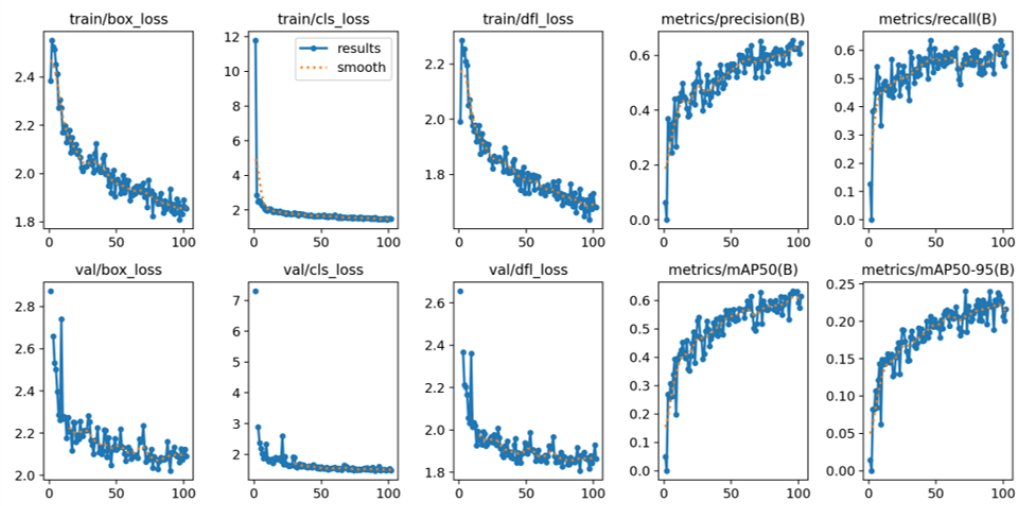


Figure C.2: YOLOv11s training curves with COCO weights

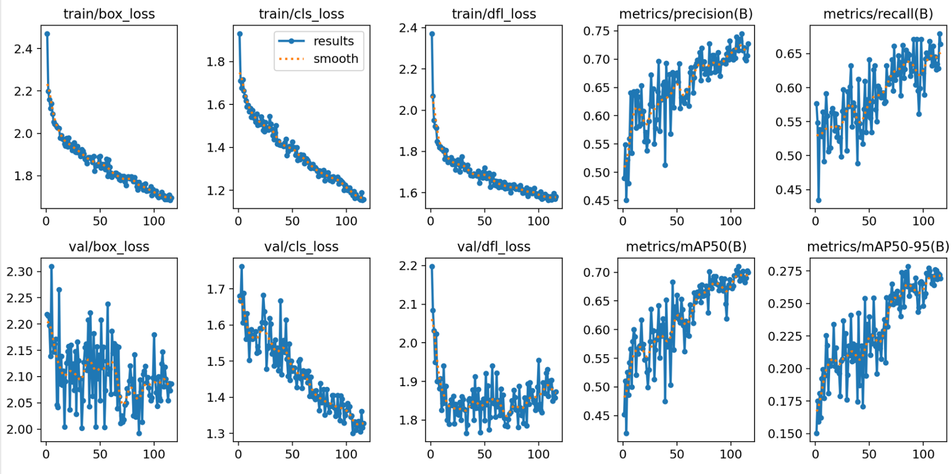


Figure C.3: YOLOv11s training curves with sentinel-2 ship dataset pretrained weight

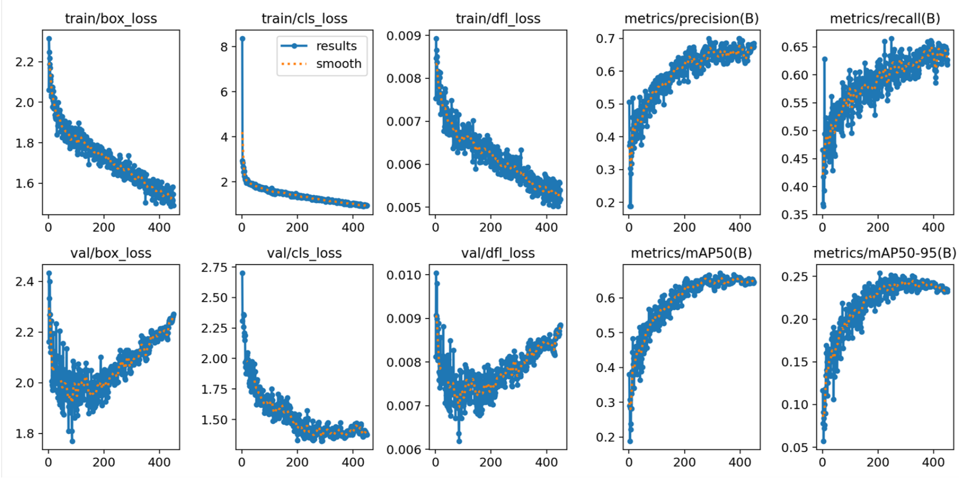


Figure C.4: YOLOv26s training curves with COCO dataset

C.1.1 Comparisons

Four YOLO-based benchmark models were evaluated in this study. YOLOv8s, YOLOv11s, and YOLOv26s were initialized with official COCO-pretrained weights before being fine-tuned on the Sentinel-2 wake detection dataset. A second YOLOv11s model was also tested using pretrained weights from a Sentinel-2 ship detection dataset containing more than 8,000 ship samples from Mäyrä et al. [2]. This benchmark was designed to compare general COCO-pretrained initialization with Sentinel-2-specific pretrained initialization for wake detection.

Metric	YOLOv8s	YOLOv11s	YOLOv11s pretrained	YOLOv26s
Precision	0.65	0.63	0.72	0.70
Recall	0.66	0.59	0.65	0.64
mAP50	0.62	0.60	0.70	0.68
mAP50–95	0.22	0.22	0.27	0.24

Table C.1: YOLO benchmark training performance

Table C.1 compares the benchmark performance of the tested models. YOLOv11s with Sentinel-2 vessel-pretrained weights achieved the best numerical results on the cropped validation dataset. However, the performance gap between this model and YOLOv26s was relatively small. It should also be noted that the YOLOv11s pretrained model had already been trained on 8,768 Sentinel-2 vessel samples before being fine-tuned on the approximately 2,000 wake samples used in this thesis. In contrast, YOLOv26s was initialized only with official COCO-pretrained weights and was then fine-tuned on the same wake dataset.

This comparison suggests that the Sentinel-2-specific pretraining provides useful prior visual features. At the same time, the relatively close performance of YOLOv26s indicates that this model still has considerable potential if Sentinel-2-specific pretraining or a larger wake dataset becomes available. In addition, qualitative inspection and full-scene testing showed that YOLOv26s performed more stably in complex scenes, such as crowded waterways, snow-covered backgrounds, and strong sea-surface clutter. For these reasons, YOLOv26s was selected as the main model for the full-scene wake detection experiments with SAHI.

Figure C.5: Comparison between YOLOv11s and YOLOv26s

Figure C.5 compares YOLOv11s with pretrained weights and YOLOv26s on the same image. The overall detection results are similar, but YOLOv26s generally produces higher confidence scores. In the upper-right part of the image, where two nearby wakes appear close to each other, YOLOv11s tends to merge them into one detection, while YOLOv26s separates them more clearly. This suggests that YOLOv26s has better local discrimination ability in crowded wake regions.

D

Appendix 4

D.1 Wake-speed Dataset Format

Table D.1: Wake-speed dataset

Field Name	**Type**	**Description**
image_name	String	Filename of the cropped GeoTIFF wake image
image_path	String	Relative path to the cropped GeoTIFF image
fixed_image_name	String	Filename of the fixed-size processed RGB wake image
fixed_image_path	String	Relative path to the processed RGB wake image
speed_kn	Float	Vessel speed over ground from AIS (knots)
speed_kmh	Float	Vessel speed converted to kilometres per hour
mmsi	Integer	Maritime Mobile Service Identity (MMSI) of matched vessel
match_dist_m	Float	Spatial matching distance between wake head and AIS position (m)
bbox_x1, bbox_y1	Integer	Upper-left pixel coordinate of wake bounding box
bbox_x2, bbox_y2	Integer	Lower-right pixel coordinate of wake bounding box
bbox_w, bbox_h	Integer	Width and height of cropped wake region in pixels
fixed_size	Integer	Final padded image size used for CNN input
use_color	Boolean	Indicates whether RGB images were used
pad_ratio	Float	Bounding-box expansion ratio applied during cropping
cx, cy	Float	Normalized YOLO bounding-box center coordinates
w, h	Float	Normalized YOLO bounding-box width and height
head_px, head_py	Float	Estimated wake-head pixel coordinates

Continued on next page

Field Name	**Type**	**Description**
head_lon, head_lat	Float	Geographic coordinates of estimated wake-head position
gray_threshold	Integer	Brightness threshold used for bright-pixel statistics
bright_pixel_count	Integer	Number of bright pixels in processed wake image
source_tiff	String	Original Sentinel-2 image tile filename
source_subdir	String	Source scene directory identifier
crs	String	Coordinate reference system of the source image

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY