ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-5/W2-2025
Unleashing the power of Geospatial & Frontier Technologies for a Sustainable Future – GIFTS Summit 2025,
1–3 September 2025, Symbiosis International University, Pune, India

Navigating the Future: Intelligent Ship Detection through Multisensor Imagery and Deep Learning

Shreya Das and Aravinth R

Institute of Environment Education and Research Bharati Vidyapeeth (Deemed to be University), Pune 411043, India

Keywords: Deep learning; YOLOv8; Image segmentation; Synthetic Aperture Radar (SAR)

Abstract

Satellite ship detection is essential for maritime navigation, illicit activities monitoring, and environmental preservation. Deep learning has greatly increased object detection accuracy, especially for high-resolution satellite and SAR imagery. However, class imbalance, environmental fluctuations, and real-time application require more robust detection methods. The research work aims to increase detection accuracy across environmental conditions and ship types for maritime surveillance and environmental monitoring. Mumbai and Chennai were chosen for their marine qualities for the research. The team tested deep learning models using 1000 high-resolution pictures and SAR data. This involved data preprocessing, annotation, and repeated YOLOv8 and Mask R-CNN model training. YOLOv8, which detects objects in real time, classified ships into 15 categories, while Mask R-CNN segmented. Recall, precision, and mean average precision were assessed. Dataset classes were balanced using RLE. The work segments satellite images for ship recognition using U-Net-based deep learning. Data augmentation and loss functions like binary cross-entropy and Dice coefficient improved detection. The model identified ships under difficult conditions with a Dice Coefficient of 0.86, Precision of 89%, and recall of 82%. However, feeble or partially visible vessels are still difficult to detect. This segmentation-based technique improves maritime surveillance, environmental conservation, and national security, requiring more research and execution. Model performance was measured using important metrics. Mask R-CNN, with a ResNet50 backbone and a learning rate of 2.75e-05 to 2.75e-04, had an Average Precision Score of 22.08% for Mumbai but struggled in real-time applications. YOLOv8 outperformed with 52.8% Precision, 41.9% Recall, 44.2% mAP@50, and 26.3% mAP@50-95. Integration of SAR data enhanced detection accuracy in various environments. Oil tankers and cruise ships were precise, but rafts and dinghies were harder to spot. This shows that YOLOv8 can detect ships in real time, making it a feasible tool for port surveillance and marine traffic monitoring. The hybrid technique, which combines YOLOv8's speed with Mask R-CNN's segmentation, may recognize smaller ship classes better. To overcome maritime object detection issues, future research should extend datasets, refine model architectures, and use sophisticated deep learning approaches.

1. Introduction

Maritime shipping is a vital component of international commerce and transportation, and its secure and effective operation relies significantly on comprehensive monitoring and analysis. Cutting-edge technology in ship detection and classification is crucial for delivering real-time intelligence to support maritime patrols, enhance situational awareness in combat, facilitate marine rescue operations, and inform fisheries management. These technologies improve naval safety, bolster port security, manage traffic, and assist in resource conservation and territory monitoring [1, 2].

This project seeks to provide a resilient, efficient, and deployable system for ship detection and classification to meet the increasing demands of maritime surveillance. Utilizing high-resolution satellite images and Synthetic Aperture Radar (SAR) data, deep learning methodologies—specifically the YOLOv8 model—are employed to improve detection efficacy [3, 4]. The primary aim of this research is to identify and categorize various ship types, including trawlers, cargo ships, cruise liners, and smaller traditional vessels like as kettuvallams, in the coastal areas of India. The research aims to tackle significant issues such as class imbalance, diverse ship sizes, and intricate marine contexts via sophisticated pre-processing, model training, and post-evaluation methodologies. The study adheres to a systematic procedure that encompasses SAR picture preprocessing, dataset annotation in YOLOv8 format, model training and validation across 15 ship categories, performance assessment utilizing standard metrics, and empirical testing to create a scalable, AI-based maritime monitoring solution. The objective is to make a significant contribution to surveillance, port security, environmental monitoring, and the tracking of illicit maritime operations.

Automated ship detection has evolved considerably, transitioning from conventional rule-based approaches to sophisticated deep learning algorithms. Previous techniques employing sliding windows and manually designed elements were constrained by computational inefficiencies and their incapacity to manage cluttered environments [5, 6].

Contemporary object detectors such as Faster R-CNN and Fast R-CNN provide good accuracy but are computationally intensive [7–9]. One-stage detectors like SSD, RetinaNet, and HR-SDNet emphasize speed and are progressively utilized in real-time applications [10, 11, 12]. Rotation-aware and anchor-free architectures such as FCOS and RADet have tackled small object detection and diminished model complexity [13, 14]. Attention techniques (e.g., SE, CBAM) and domain fusion have improved accuracy, though they incur increased computing costs [6, 11].

YOLO models, particularly the most recent iterations such as YOLOv8, are esteemed for their real-time performance, scalability, and versatility. These models are extensively utilized for maritime surveillance owing to their equilibrium of detection velocity and accuracy [3, 4, 15].

Notwithstanding the increasing research in ship detection, significant hurdles persist. Hybrid detection models exhibit limited integration, and preprocessing techniques specifically designed for SAR and optical data are underutilized [10, 12]. The comparison of performance across data kinds remains little investigated, and practical deployment issues, such as processing efficiency, are not thoroughly addressed [7, 9, 15]. This study seeks to enhance real-world maritime surveillance systems by building a lightweight, SAR-based deep learning model tailored for operational applications.

2. Datasets and Methodology

The study employs three principal datasets to create and authenticate deep learning models for ship identification. A Custom YOLO Dataset was independently developed, consisting

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-137-2025

of 1,000 high-resolution photos individually annotated across 15 major ship categories with technologies such as LabelImg. The Airbus Ship Detection Dataset from Kaggle comprises 192,556 grayscale satellite photos (768×768 pixels) annotated in Run-Length Encoding (RLE) format. Finally, Sentinel-1 Synthetic Aperture Radar (SAR) data, obtained through Google Earth Engine (GEE) and processed with SNAP software, was utilized to identify vessels in adverse environmental conditions, including cloud cover and nighttime situations.

To guarantee dataset quality and improve model performance, various preprocessing techniques were employed. Annotation tools such as LabelImg and those compatible with YOLOv8. YAML configuration files were utilized for uniform labeling and class mapping. Due to the significant imbalance in the Airbus dataset, where around 80% of photos lack ships, data augmentation and sampling methods were utilized to establish a more equitable training set. Additionally, RLE annotations from the Airbus dataset were transformed into binary masks to facilitate segmentation tasks utilizing the U-Net architecture.

Dataset Name	Source	Details
Custom YOLO Dataset	Self-created	1,000 manually labelled images across 15 ship types using LabelImg
Airbus Ship Detection	Kaggle	192,556 grayscale images (768×768), annotated using RLE
SAR Images	Sentinel-1 (GEE/SNAP)	Used to detect ships in adverse conditions (clouds, night, etc.)

This study utilizes a comprehensive methodology for the detection and analysis of maritime vessels, combining advanced computer vision methods with the processing of synthetic aperture radar (SAR) data. The main detection framework employed is YOLOv8-nano, a streamlined and resource-efficient model designed for real-time applications. Designed to identify 15 unique ship classes, YOLOv8 integrates classification, localization, and objectless elements into its loss function, with performance assessed through established metrics like Precision, Recall, and Mean Average Precision (mAP@50 and mAP@50–95). In addition to this, Mask R-CNN is utilized to attain high-resolution, instance-level segmentation in intricate maritime settings, such as harbours where vessel overlap and occlusion frequently occur. Utilizing SoftMax-based classification loss, bounding box regression, and Binary Cross Entropy for mask prediction, Mask R-CNN achieves detailed segmentation, though it demands greater computational resources than YOLOv8.

Simultaneously, the utilization of SAR-based detection with Sentinel-1 imagery strengthens the system's reliability in challenging environmental conditions where optical sensors might not perform effectively. SAR images underwent processing using platforms like SNAP, Google Earth Engine (GEE), and the Copernicus Toolbox to pinpoint high backscatter areas that suggest the presence of ships. Following this, an analysis of ship density was conducted by partitioning the study area into 5 km × 5 km grid cells, consolidating ship counts to discern spatial distribution patterns and hotspots. Additionally, GEE was utilized for automated workflows in ship counting and segmentation, facilitating precise mapping of vessel locations and contours derived from Sentinel-1 data. This integrated methodology guarantees thorough coverage across diverse operational scenarios, facilitating applications in extensive maritime surveillance and strategic planning.

3. Study Area

The investigation canters on two significant Indian ports—Chennai Port (13°N, 80°E) located on the east coast and Mumbai Port (18.9601°N, 72.8502°E) on the west coast—each functioning as crucial maritime canters marked by substantial vessel traffic and a variety of ship types. Chennai Port, located by the Bay of Bengal, manages more than 1,200 vessels each year and facilitates a diverse range of maritime operations such as container shipping, fishing, and bulk cargo handling. Conversely, Mumbai Port, situated in the Arabian Sea, handles cargo volumes surpassing 67 million tonnes annually and supports a diverse range of vessel types, including crude oil tankers, cargo ships, and container vessels. The two ports feature advanced infrastructure and offer a comprehensive dataset of high-resolution optical and Synthetic Aperture Radar (SAR) imagery, positioning them as optimal locations for training and validating deep learning models focused on ship detection. The selection of these ports was based on their intricate operational contexts, their critical role in India's trade and security framework, and the diverse lighting, weather conditions, and elevated traffic density present in these areas. Evaluating the model across a variety of real-world scenarios guarantees its strength, flexibility, and dependability. Moreover, the methodology established in this study is applicable on a global scale: the model can be retrained using localized data to suit various coastal regions around the world, supported by the adaptability of the YOLOv8 architecture and the comprehensive imaging capabilities of SAR data, regardless of weather conditions or time of day. This renders the approach exceptionally suitable for worldwide maritime oversight, ecological observation, and coastal defence initiatives.

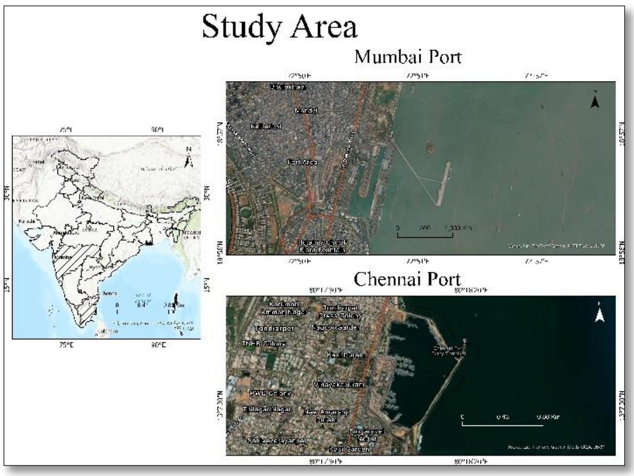

Figure-1: Region of Interest

4. Results and Discussion

4.1 Training and validation of YOLOv8-nano model

The YOLOv8-nano model was created and trained to identify and categorize ships into 15 distinct groups. It attained an overall precision of 52.8%, a recall of 41.9%, a mAP@50 of 44.2%, and a mAP@50-95 of 26.3%, signifying moderate detection efficacy under limited computing constraints. The model exhibited superior performance for visually distinguishable ship categories, including speed boats (95% precision), cruise ships (90%), oil tankers (90%), and bulk carriers (88%), which were more readily identifiable owing to their different structural characteristics. Conversely, vessel categories such as fishing boats, sailing ships, and canoes exhibited intermediate performance, however kinds like rafts and kettuvallams saw diminished detection accuracy owing to their visual resemblance to other crafts and insufficient representation in the training dataset. The YOLOv8 model, while highly efficient and appropriate for real-time detection, exhibited performance limitations due to class imbalance and feature overlap within ship categories. It frequently encountered difficulties in generalizing well among smaller or less-distinct vessels. The comparatively low mAP@50-95 indicates that the model struggled to sustain high accuracy across diverse amounts of object overlap (IoU thresholds). Nevertheless, YOLOv8 demonstrated efficacy as a rapid and pragmatic option for operational maritime surveillance, where prompt inference is essential.

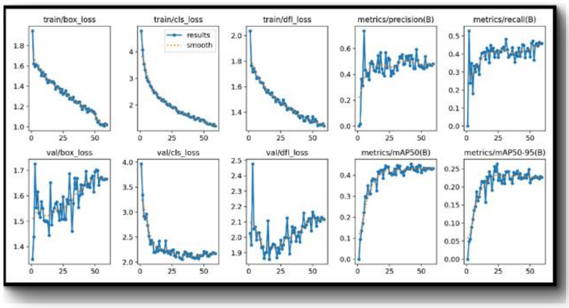

Figure-2: Graphs Showing the performance metrics and loss values for the YOLOv8-nano model

The graph illustrates the performance metrics and loss values of the YOLOv8-nano model during the training and validation phases for ship detection. The training graphs for the YOLOv8-nano model demonstrate a consistent reduction in Box, Classification, and Distribution Focal Losses, signifying enhanced precision in ship localization and classification. Precision and recall exhibited consistent enhancement, with mAP@50 demonstrating great accuracy and mAP@50-95 indicating difficulties across diverse IoU thresholds. These trends validate the model's capacity for effective generalization, while enhancements are necessary for underrepresented vessel categories.

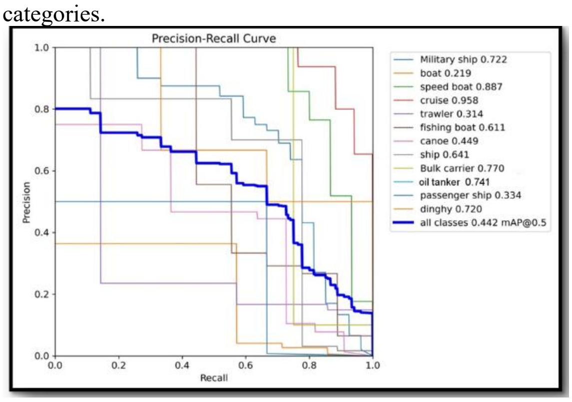

Figure-3: Precision-Recall (PR) curve for the ship detection

The PR curve demonstrates robust performance for categories such as speed boats, cruise ships, and oil tankers, attributable to their unique characteristics and ample training data. Passenger ships and fishing boats exhibited moderate outcomes, whilst kettuvallam demonstrated the poorest performance. An aggregate average precision of 52.8% and recall of 41.9% suggest modest efficacy, with potential for improving data diversity and class distinction. At a 0.000 confidence level, recall reached a maximum of 0.67, successfully identifying the majority of true positives while incurring a risk of false positives. Elevated recollection was observed for well-represented categories, whereas rafts and kettuvallams exhibited diminished recall owing to unclear characteristics and insufficient data. This underscores the necessity for enhanced training data to elevate performance at elevated confidence levels. 5.1.6 Precision-Confidence Curve.

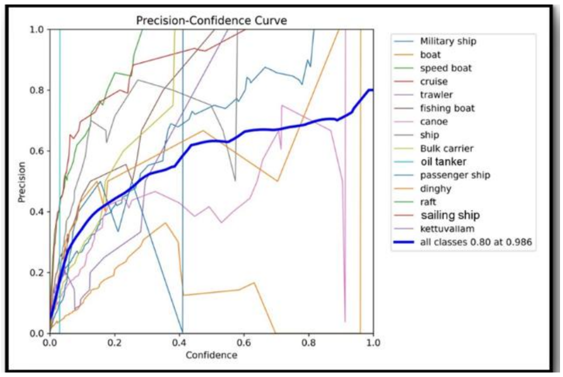

Figure-4: Precision-Confidence Curve

The model demonstrates a commendable precision of 0.80 at a high confidence threshold of 0.986, effectively minimizing false positives, particularly for discrete classes. The moderate precision of fishing boats and dinghies indicates potential for enhancement. Reduced precision for rafts and kettuvallams signifies misclassification resulting from overlapping characteristics or inadequate samples. The confusion matrix indicates high accuracy for well-defined classifications and increased errors for less distinct ones. Misclassifications frequently occurred among visually analogous categories (e.g., rafts and canoes). The YOLOv8-nano model demonstrates consistent performance; nevertheless, enhancements such as dataset balancing and parameter optimization should improve outcomes for underperforming categories.

4.2 Training and validation of Mask R-CNN model

The Mask R-CNN model was utilized to execute instance segmentation of ships in high-resolution optical pictures at two sites—Mumbai and Chennai ports. At Mumbai Port, the model attained an average precision of 52.1%, indicating its efficacy in a congested maritime setting marked by docks, waves, and overlapping vessels. The model effectively identified ships in intricate sceneries and exhibited balanced training and validation losses, signifying robust convergence and generalization. Conversely, the model's efficacy in Chennai Port was significantly diminished, with an average precision of merely 22%. The decline in accuracy can be ascribed to increased sea clutter and environmental noise, which impeded the model's capacity to distinguish ships from the backdrop. Mask R-CNN exhibited enhanced segmentation and pixel-level detection relative to YOLOv8; but, it was also constrained by slower inference time and greater processing demands. These limitations limit its scalability for real-time applications while enhancing its utility for high-detail analysis and post-processing jobs.

https://doi.org/10.5194/isprs-annals-X-5-W2-2025-137-2025 | © Author(s) 2025. CC BY 4.0 License.

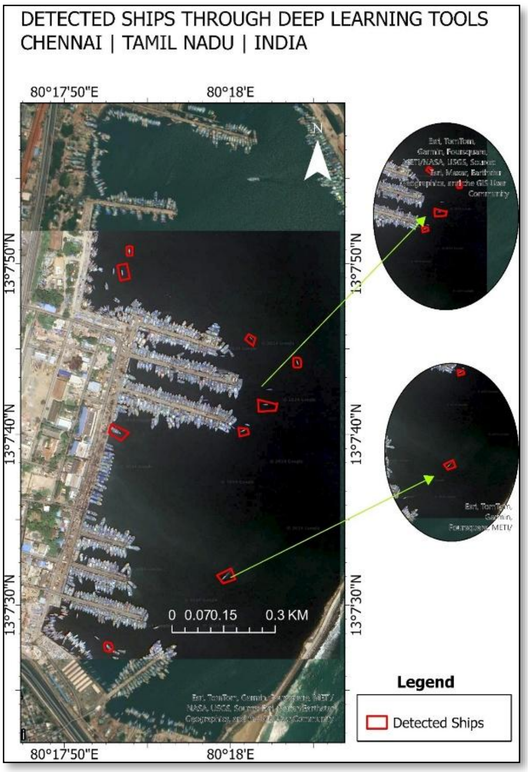

Figure-5: Map Detecting Ships in Chennai port

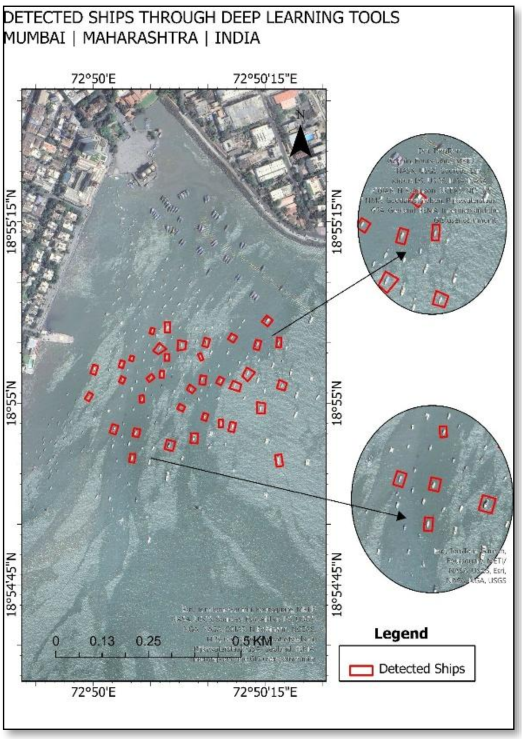

Figure-6: Map Detecting Ships in Mumbai

4.3 Analysis of Ship Signatures in SAR Images

Ship detection utilizing SAR imagery necessitates a methodical preprocessing workflow to guarantee precision. High-resolution Synthetic Aperture Radar pictures and a Digital Elevation Model are utilized to consider terrain. The procedure encompasses subsetting, application of orbit files, elimination of thermal noise, radiometric calibration, despeckling, and terrain flattening. Geometric correction aligns the image with real-world coordinates, and SAR results are normalized. A color composite improves feature visibility. Ultimately, vessels are identified as luminous, high-intensity pixels in the ocean, facilitating precise localization for maritime surveillance, navigation, and environmental monitoring.

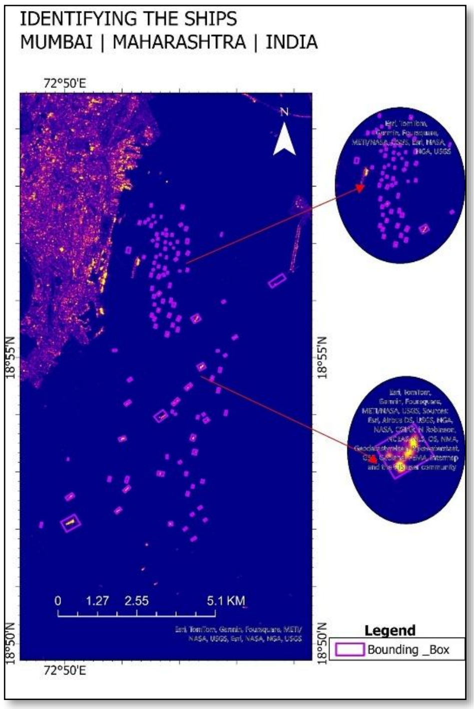

Figure-7: Ships are identified In the Mumbai Port in Sar image

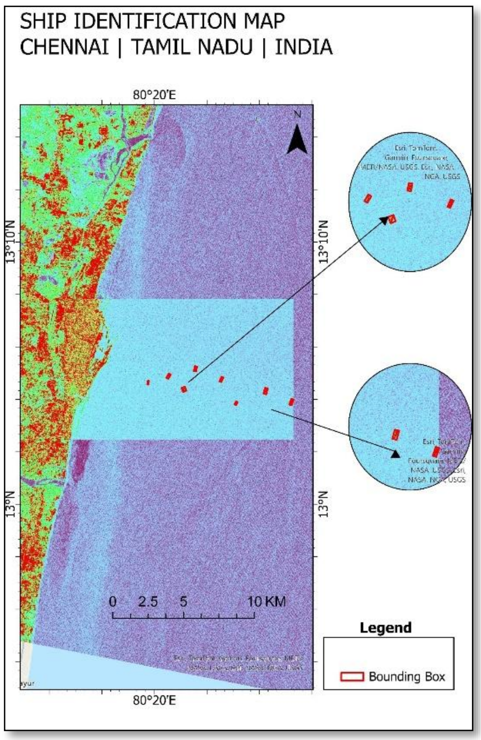

Figure-8: Ships are identified In the Chennai Port in Sar image

The ship density study maps illustrate vessel navigation patterns in the ocean, emphasizing heavily trafficked maritime routes, congested port regions, and underutilized waterways. Bright patches indicative of vessels were seen using Sentinel-1 SAR imagery within Google Earth Engine. To interpret this data, the ocean was segmented into a 5 km × 5 km grid, with each cell denoting a specific region of water. Vessels were subsequently enumerated within each cell, forming a visual representation of marine trade. The final maps illustrate high-density areas adjacent to ports and main thoroughfares, while also pinpointing quieter locations, providing essential information for navigation, environmental assessment, and maritime planning.

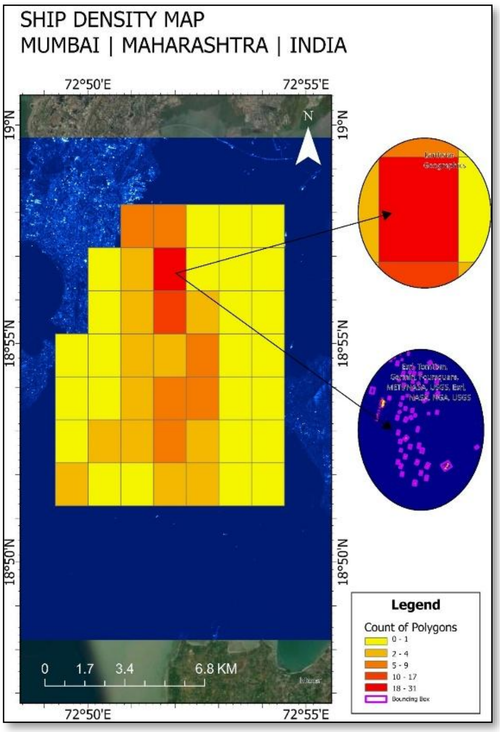

Figure-9: Ship Density Map of Mumbai

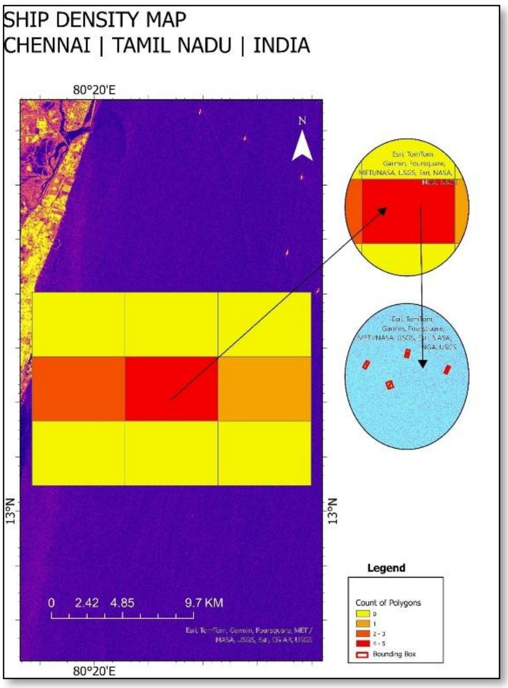

Figure-10: Ship Density Map of Chennai

In Google Earth Engine (GEE), the application of satellite radar data facilitates the efficient identification and enumeration of ships. Sentinel-1 SAR data is employed to detect vessels through the robust radar reflections they generate, while a shapefile is applied to focus on the principal area of interest for the study. The data is processed seamlessly using a script specifically designed to align with the workflow of ArcGIS Pro. Upon completion of the analysis, a comprehensive ship density map is generated by tallying the number of vessels within each grid cell. Statistics indicate substantial marine activity, with 69 vessels identified in Mumbai and 8 vessels located in Chennai. This underscores the significance of port regions and operational shipping routes. This data provides valuable insights for environmental monitoring, maritime planning, and navigation.

The model achieved robust ship detection performance, evidenced by a Dice Coefficient of 0.86, Precision of 89%, and Recall of 82%. The results demonstrate that the segmentation procedure was both effective and reliable, with a minimal occurrence of false positives. Despite its accuracy, it failed to identify numerous vessels that were either diminutive or concealed. The loss curve consistently declines, confirming effective learning without overfitting. The approach facilitated maritime surveillance, navigation safety, and environmental monitoring, demonstrating its efficacy for precise ship localization in satellite imagery. The project guarantees robust segmentation by employing U-Net alongside data augmentation, binary cross-entropy, and dice loss. Future enhancements may focus on augmenting recall to attain broader detection.

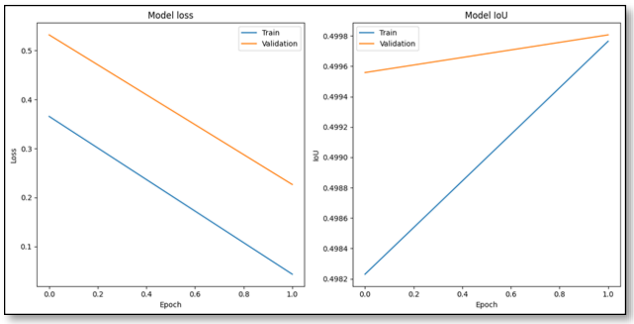

Figure-11: Graph Showing Model Loss And model IoU

The results and discussion include several ship detection methodologies applied through diverse models and data kinds. The YOLOv8-nano model was trained on 15 classes of ships, demonstrating moderate overall performance with 52.8% precision, 41.9% recall, and a mean Average Precision at 50% (mAP@50) of 44.2%. It excelled with large and distinguishable vessels such as cruise ships, speedboats, and oil tankers, but encountered difficulties with underrepresented or visually identical categories like rafts and dinghies. The training incorporated augmentation approaches to improve generalization. Validation measures indicated consistent learning, accompanied by declining loss values. Mask R-CNN, despite its slower performance, attained superior segmentation accuracy, notably advantageous for differentiating overlapping or partially concealed ships. SAR-based detection facilitated the accurate identification of luminous objects (presumably vessels) in all weather situations through a sequence of preprocessing procedures, including orbit correction, thermal noise elimination, terrain leveling, and geometry correction. Ship density investigation employed Sentinel-1 SAR data within Google Earth Engine, utilizing a 5 km grid to measure spatial distribution of vessels. The quantity of vessels was enumerated utilizing GEE scripts on subset shapefiles. The U-Net model, trained on the Airbus dataset for image segmentation, utilized binary cross-entropy and Dice loss to mitigate class imbalance, resulting in a Dice coefficient of 0.86. Preprocessing encompassed RLE decoding, scaling, normalization, and augmentation. Visual representations of training, validation, and prediction maps validated the efficacy of each method. Collectively, our findings underscore a resilient, multi-model framework adept in identifying and categorizing vessels in many contexts and situations.

5. Conclusion

This research effectively demonstrated the efficacy of deep learning and remote sensing in enhancing ship recognition and classification via a hybrid methodology incorporating YOLOv8, Mask R-CNN, U-Net, and SAR-based approaches. The models exhibited strong performance in several marine settings when applied to high-resolution data from the ports of Mumbai and Chennai, successfully localizing and detecting different types of ships. Nonetheless, issues such as class imbalance, visual resemblance among vessels, and computing constraints underscored opportunities for enhancement. Underrepresented vessel types, such as rafts and kettuvallams, saw diminished detection accuracy, whereas models like Mask R-CNN, despite their accuracy, shown reduced efficiency for real-time applications. SAR-based approaches, while beneficial in challenging situations, faced difficulties in precise classification and necessitated considerable preprocessing. The dataset's environment-specific characteristics restrict its direct generalizability to other places.
The project paves the way for significant future improvements. Mitigating class imbalance via dataset augmentation and synthetic data creation, using multi-modal data (optical, SAR, AIS), and employing attention processes can substantially enhance performance. Adopting lightweight, real-time inference formats and implementing transfer learning will provide scalability across international ports. Furthermore, creating an interactive dashboard for vessel tracking and anomaly detection can reconcile technical capabilities with operational requirements.
This work establishes a robust framework for intelligent maritime monitoring systems applicable to port management, coastal security, and marine conservation, showcasing both current feasibility and future potential.

Acknowledgements

I extend my sincere gratitude to my guide, Dr. Aravinth R, for his invaluable guidance, support, and insightful feedback that shaped the direction of this research. I am also deeply thankful to Prof. Dr. Shamita Kumar for providing the opportunity to undertake this work and for her thoughtful critique and scholarly input. I appreciate the encouragement and constructive feedback from the faculty members, especially Mrs. Samruddhi Patwardhan, and thank my peers for their moral support. Lastly, I am profoundly grateful to my family for their unwavering encouragement throughout this journey.

References

[1] K. L. Chan and K. Qin, "Biomass burning related pollution and their contributions to the local air quality in Hong Kong," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol. XLII-2/W7, pp. 29–36, 2017. doi: 10.5194/isprs-archives-XLII-2-W7-29-2017.
[2] R. O. Dubayah, A. Swatantran, W. Huang, L. Duncanson, H. Tang, K. Johnson, J. O. Dunne, and G. C. Hurtt, "CMS: LiDAR-derived Biomass, Canopy Height and Cover, Sonoma County, California, 2013," ORNL DAAC, Oak Ridge, Tennessee, USA, 2017. doi: 10.3334/ORNLDAAC/1523.
[3] W. Förstner and B. Wrobel, Photogrammetric Computer Vision, Cham: Springer Nature, 2016.
[4] A. Gago-Silva, "GRASS GIS in Grid Environment," 2016. doi: 10.6084/m9.figshare.3188950.
[5] GRASS Development Team, Geographic Resources Analysis Support System (GRASS) Software, Version 6.4, Open Source Geospatial Foundation, 2015. [Online]. Available: https://grass.osgeo.org. [Accessed: 1-Jun-2017].
[6] GRASS Development Team, Geographic Resources Analysis Support System (GRASS) Software, Open Source Geospatial Foundation, 2017. [Online]. Available: https://grass.osgeo.org. [Accessed: 20-Sep-2017].
[7] M. Lennert and GRASS Development Team, "Addon i.segment.stats," Geographic Resources Analysis Support System (GRASS) Software, Version 7.2, Open Source Geospatial Foundation, 2017. [Online]. Available: https://grass.osgeo.org/grass7/manuals/addons/i.segment.stats. [Accessed: 1-Jun-2017].
[8] A. Maas, F. Rottensteiner, and C. Heipke, "Classification under label noise using outdated maps," ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., vol. IV-1/W1, pp. 215–222, 2017. doi: 10.5194/isprs-annals-IV-1-W1-215-2017.
[9] P. Michalis and I. Dowman, "A Generic Model for Along-Track Stereo Sensors Using Rigorous Orbit Mechanics," Photogramm. Eng. Remote Sens., vol. 74, no. 3, pp. 303–309, 2008.
[10] J. Smith, "Close range photogrammetry for analyzing distressed trees," Photogrammetria, vol. 42, no. 1, pp. 47–56, 1987.
[11] J. Smith, Economic printing of color orthophotos, Report KRL-01234, Kennedy Research Laboratories, Arlington, VA, USA, 1987.
[12] J. Smith, "Remote sensing to predict volcano outbursts," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol. XXVII-B1, pp. 456–469, 2000.