Using R-CNN for Ship Detection in SAR and EO Imagery

Mason Palmer and Brad Jensen

***Abstract*— We combine Synthetic Aperture Radar (SAR) and Electro-Optical (EO) imagery to develop two separate ocean vessel detection systems. SAR imagery excels in all-weather, day-and-night surveillance, while EO imagery captures detailed visual features, enabling complementary data integration. Utilizing state-of-the-art models, such as Faster R-CNN and YOLO, we optimize detection accuracy across diverse environmental conditions.**

Our evaluation shows that while the SAR-based model underperforms the challenge baseline due to computational limitations, the EO-based YOLO models achieve promising results. Class-level analysis highlights challenges in detecting certain vessel types, such as motorboats and fishing vessels, underscoring the need for further refinements. Finally, we discuss the potential of future advancements, including the integration of AIS data, multimodal fusion, and temporal analysis, to enhance real-time monitoring and predictive capabilities for maritime security applications. This work lays a foundation for leveraging machine learning and satellite imagery to address the pressing issue of shadow fleet detection.

Keywords—SAR, YOLO, detection, computer vision, Artificial Nerual Networks, Machine Learning

I. Introduction

Detecting shadow fleets—vessels that deactivate or misuse AIS to evade detection—addresses a critical real-world challenge with far-reaching economic, environmental, and societal impacts. These fleets often engage in illicit activities such as smuggling, human trafficking, piracy, and illegal fishing, creating hazards like increased collision risks and unfair competition. This project focuses on using machine learning and multimodal satellite imagery to detect ships and analyze their behavior, providing tools to combat these issues.

We combined Synthetic Aperture Radar (SAR) and Electro-Optical (EO) imagery to develop a shadow fleet detection system. SAR imagery is uniquely suited for maritime surveillance due to its consistent performance in all weather and lighting conditions, while EO imagery excels in capturing visual details like color and patterns. Using models like Faster R-CNN and YOLO, we aimed to maximize detection accuracy across varied environmental conditions.

This report outlines our methods, results, and the potential for future advancements, including real-time monitoring and predictive capabilities to enhance maritime surveillance and combat the threats posed by shadow fleets.

II. Related Work

Understanding and addressing shadow fleets builds upon key insights and methodologies from prior research. One contribution is the Fighting Shadows and Ghosts framework, which introduces Identification Deception Tactics (IDTs) and the MARINE approach. This framework outlines shadow fleet operations, showing many tactics used to evade detection and offering a structured methodology for identifying these covert operations.

Another essential study ranks ship detection methods using SAR imagery, using machine learning and artificial intelligence to establish model evaluation techniques and performance benchmarks. This provides valuable guidance for selecting and tuning detection algorithms.

Further advancements in ship monitoring come from research into velocity estimation in SAR imagery by using multitask deep learning. This work demonstrates the potential of incorporating velocity as an input parameter, paving the way for predictive modeling and anomaly detection in maritime applications.

By building on these foundational studies, this project and future projects can integrate proven methodologies and innovative approaches to enhance shadow fleet detection and analysis. These related works give a base to the feasibility and potential impact of combining SAR and EO imagery with machine learning for maritime surveillance.

III. Proposed Methodology

A. Data Preprocessing

To prepare the data for model training and analysis, several preprocessing steps were implemented to optimize the quality and usability of the dataset.

For the SAR data, we applied a Lee filter to despeckle the imagery. The Lee filter was selected for its simplicity, speed, and effectiveness in preserving edges—qualities that have made it an industry standard in SAR data processing. Following despeckling, the VV polarization and VH polarization were combined into a single two-channel image. Since the base model requires three channels, we filled the third channel with the average of the VV and VH values, ensuring compatibility with the model while preserving meaningful data.

Both the EO and SAR images were then split into smaller regions, a crucial step for improving the model's performance. By creating smaller image patches, the model was better able to focus on localized targets, such as small maritime vessels, that might have been missed in larger scenes. This approach also reduced the complexity of the background within each patch, allowing the model to more effectively learn and converge on the relevant features.

To handle the corresponding labels during this process, a custom script was developed to reassign bounding box values accurately, adhering to the formatting requirements of the respective datasets: YOLO .txt files for EO data and COCO.geojson files for SAR data. Additionally, multiple patch sizes were generated, both to ensure that images truncated during the splitting process could still be included in a complete patch and to provide an added layer of data augmentation. This approach not only enriched the dataset but also improved the model's robustness by exposing it to a wider variety of localized features and configurations.

Please see our project Github repository at https://github.com/mason-palmer/cs-5640 to see code details.

B. SAR Model Creation and Validation

The machine learning framework Detectron2, developed by Meta (formerly Facebook), was employed to facilitate the creation and training of the object detection model. Detectron2 provides a robust and flexible platform for implementing state-of-the-art deep learning models, enabling seamless integration with the unique requirements of SAR data processing. A pre-trained base model provided by Detectron2 was used as a foundation, leveraging transfer learning to accelerate development and improve performance on the maritime vessel detection task.

Several specific settings and hyperparameters were selected to optimize the model for detecting small maritime vessels in SAR imagery:

- Normalization: The pixel mean was set to 123 and the standard deviation to 58 for each respective channel. This normalization ensured compatibility with the pretrained weights, facilitating effective transfer learning while preserving the unique characteristics of SAR data.

- Anchor Sizes: Smaller anchor sizes [8,16,32,64,128][8, 16, 32, 64, 128][8,16,32,64,128] were chosen to better capture small maritime vessels that dominate the dataset. Larger anchor sizes were avoided to prevent overlooking these smaller targets, which are critical for the detection task.

- Automatic Mixed Precision (AMP): AMP was enabled to improve computational efficiency and reduce memory usage during training. This optimization allowed the model to handle the high-resolution SAR data effectively on available hardware.

- Score Threshold for Predictions: The score threshold for predictions was set to 0.5 to ensure that only high-confidence detections were considered. This choice reduced the likelihood of false positives, thereby improving the precision of the model.

- Training Checkpoints: Checkpoints were saved every 50 iterations, enabling regular evaluation of model performance during training. The best-performing checkpoints were identified based on metric outputs and subsequently used for continued training, with settings adjusted iteratively to refine the model further.

This configuration ensured that the model was both computationally efficient and tailored to the unique challenges posed by SAR imagery, resulting in a highly specialized framework for maritime vessel detection.

C. EO Model Creation and Validation

To optimize the detection of maritime vessels in electro-optical (EO) imagery, various YOLO model architectures and sizes were explored, including YOLOv8n, YOLOv8s, YOLOv11n, and YOLOv11s. The dataset was split into 80% training, 10% validation, and 10% testing subsets to ensure proper evaluation and prevent overfitting. A range of data augmentation techniques was applied to enhance the robustness of the models. These included hue and brightness jitter to simulate varying lighting conditions, geometric transformations such as rotations and flips to account for different orientations, and advanced techniques like mosaic and cutout, which provided additional context and variability for training. Performance evaluation was guided by F1 score and mAP-50 metrics, ensuring a balanced focus on precision and recall while driving iterative model improvement.

IV. Experiments

A. SAR Dataset

The xView3 dataset was developed as part of a data science challenge aimed at improving the detection of boats to support dark ship classification. This initiative was jointly sponsored by the Defense Innovation Unit (DIU)—a Department of Defense organization dedicated to fostering technological innovation—and Global Fishing Watch, a nonprofit committed to promoting the monitoring of ocean activities to combat illegal fishing and other unlawful maritime practices. The dataset is designed to provide a foundation for advancing machine learning methods that enhance maritime domain awareness.

Each scene in the dataset consists of two Synthetic Aperture Radar (SAR) images captured in VV (vertical transmit, vertical receive) and VH (vertical transmit, horizontal receive) polarizations, providing diverse data characteristics for analysis. The images are georeferenced using the Universal Transverse Mercator (UTM) projection and have a spatial resolution of 500 meters, ensuring high precision and consistency for geospatial tasks. These SAR images were collected by the European Space

Agency's Copernicus Sentinel-1 satellites, which are renowned for their reliability and advanced capabilities in remote sensing.

The dataset primarily includes scenes of shoreline and near-shore areas, avoiding deep-sea regions, to emphasize the detection of vessels operating within these high-activity zones. This focus aligns with the challenge's objective of identifying and tracking vessels in environments where illegal or untracked maritime activities are more prevalent. The xView3 dataset represents a critical step toward leveraging SAR imagery for impactful applications in maritime surveillance and security.

B. EO Dataset

The xView EO Aerial Dataset was developed as part of a data science challenge aimed at enhancing aerial classification models for applications in disaster recovery and response. Originally designed to include a wide range of classifications, such as various types of vehicles and infrastructure, the dataset offered a comprehensive foundation for improving machine learning models in diverse contexts.

For this specific adaptation, the dataset was refined to focus exclusively on maritime vessels. This involved removing all images that did not contain boats, streamlining the dataset to address the unique challenges associated with vessel detection and classification. By narrowing the scope to maritime contexts, the adapted dataset provides a targeted resource for developing and evaluating models aimed at improving vessel identification, particularly in situations where accurate maritime classification is critical. This tailored approach supports efforts to optimize disaster response and enhance situational awareness in maritime environments.

V. Baseline Description

A. SAR Baseline

The xView3 challenge employed the F1 score as the primary evaluation metric to assess the effectiveness of models in detecting maritime vessels in SAR imagery. A baseline model was provided as a point of comparison for participants, offering a benchmark against which improvements could be measured. This baseline model facilitated consistency and allowed participants to gauge their performance relative to a standardized approach. The results of the competition highlighted the potential for improvement, with the winning models achieving significantly higher F1 scores, demonstrating the effectiveness of their novel techniques and optimizations. These scores provide valuable context for understanding the advancements made through this study's SAR model configurations and preprocessing strategies.

B. EO Baseline

For the EO dataset, multiple YOLO versions and sizes were compared to establish a performance baseline. Each YOLO version introduces unique architectural changes to its backbone, with specific versions excelling in different tasks depending on the dataset and target characteristics. Versions such as YOLOv8n, YOLOv8s, YOLOv11n, and YOLOv11s were systematically evaluated to determine their effectiveness for maritime vessel detection. The F1 score, a balanced metric combining precision and recall, was used to measure performance consistently across these configurations. By analyzing the F1 scores of each YOLO version and size, this study identified models that achieved a strong trade-off between detection accuracy and computational efficiency.

Additionally, the comparison of model sizes highlighted the importance of finding an optimal balance between smaller models, which are computationally lightweight, and larger models, which offer higher accuracy but demand more resources. This analysis informed the selection of a configuration that maximized performance while maintaining practical resource requirements for maritime detection tasks. The incorporation of F1 score metrics ensured a rigorous and standardized evaluation process, guiding the refinement of the EO baseline and the final model selection.

VI. Experimental Evaluation

A. SAR Model Results

SAR Model Performance		
Model	Vessel Classification F1 Score	Fishing Classification F1 Score
xView3 reference Model	0.430	0.395
Our Model	0.362	0.233

The table compares the performance of the SAR-based models in terms of vessel classification and fishing vessel classification F1 scores. The xView3 reference model achieves higher F1 scores for both tasks, with 0.430 for vessel classification and 0.395 for fishing vessel classification. In contrast, the custom model developed in this project achieves F1 scores of 0.362 and 0.233, respectively, for these tasks.

These results indicate that the custom model underperforms compared to the reference model. The lower scores likely stem from factors such as limited training iterations, computational constraints, and possibly insufficient hyperparameter tuning or preprocessing optimizations. The significant gap in performance, particularly for fishing vessel classification, suggests that the custom model may struggle with distinguishing smaller or less distinctive vessel types in SAR imagery. Further training, improved data preprocessing, and targeted adjustments to the model architecture could help close this performance gap.

B. EO Model Results

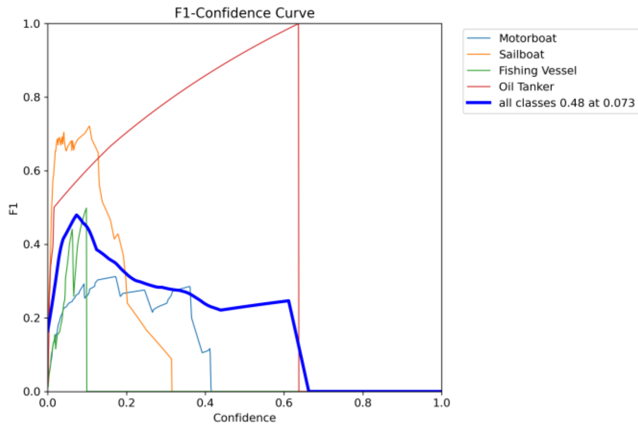

Fig. 1. YOLOv8n F1-Confidence Curve

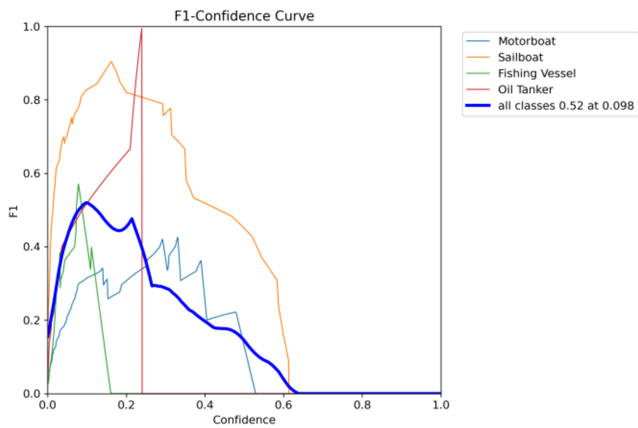

Fig. 2. YOLOv8s F1-Confidence Curve

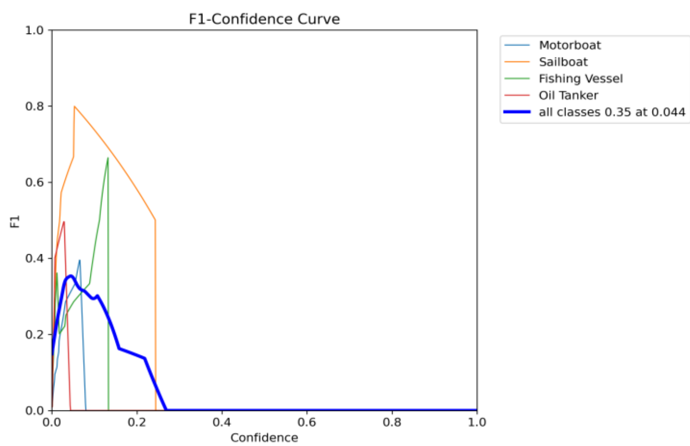

Fig. 3. YOLO11n F1-Confidence Curve

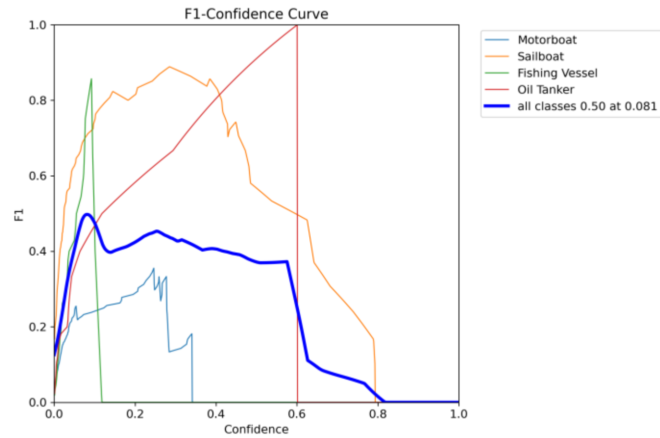

Fig. 4. YOLO11s F1-Confidence Curve

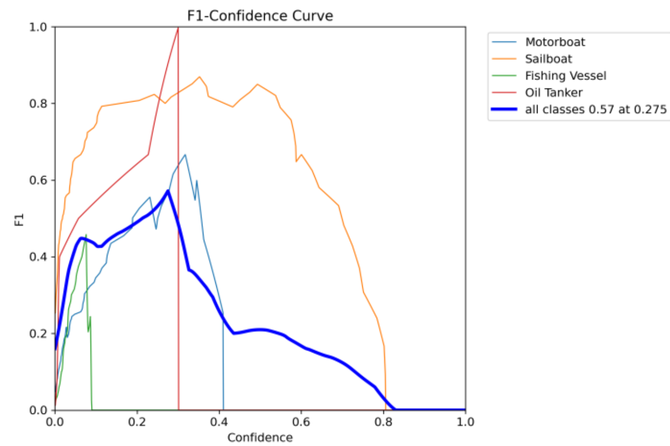

Fig. 5. YOLO11m F1-Confidence Curve

The comparison of the F1-confidence curves across the YOLO models reveals several insights regarding the performance differences between architectures (YOLOv8 vs. YOLOv11) and model sizes (n, s, and m).

YOLOv8 models generally achieve lower peak F1 scores compared to YOLOv11 models, with YOLOv8s reaching a peak F1 score of 0.52 at a confidence threshold of 0.098, while YOLOv11s achieves 0.50 at 0.081. YOLOv11m demonstrates the highest peak F1 score of 0.57 at a confidence threshold of 0.275, indicating its superior accuracy. Narrow models, such as YOLOv8n and YOLOv11n, underperform significantly, achieving lower peak F1 scores of 0.48 and 0.35, respectively, and displaying sharp performance declines at higher confidence thresholds. These results suggest that larger models, like YOLOv11m, benefit from additional capacity to capture complex features in the data, whereas smaller models, like YOLOv8s, balance accuracy and computational efficiency more effectively.

VII. CONCLUSION

A. Findings

For the SAR data, we were unable to match the F1 scores of the challenge baseline. This outcome may be attributed to an underestimation of the time and computational resources necessary to fully develop a stable and high-performing model for this data. Analysis of the training logs indicates that the model was still in the early stages of convergence, with insufficient training iterations to achieve optimal performance.

Several factors contributed to these limitations. The complexity of the SAR data, combined with the computational demands of processing it effectively, required resources beyond what was available during this project. Specifically, a more powerful GPU and additional time for training would likely allow the model to reach its full potential.

Despite these challenges, we remain confident in our methodology. With more time to fine-tune hyperparameters, explore advanced data augmentation techniques, and utilize state-of-the-art hardware, our approach has the potential to significantly improve performance and achieve F1 scores closer to or exceeding the challenge baseline.

In regard to EO data, the results highlight a clear tradeoff between model size and performance. YOLOv11m demonstrates the best overall detection accuracy, making it ideal for applications where computational resources are not a limiting factor. YOLOv8s, while slightly less accurate, provides a strong balance of efficiency and performance, making it suitable for resource-constrained environments. Narrow models, such as YOLOv8n and YOLOv11n, are less effective, with lower peak F1 scores and reduced robustness, limiting their applicability in complex detection scenarios.

These findings emphasize the importance of selecting the appropriate model architecture and size based on the computational constraints and performance requirements of the application. Additionally, addressing the performance gap for underperforming classes such as motorboats and fishing vessels could further enhance the detection capabilities of these models.

B. Future Work

There are numerous opportunities for extending this work, particularly through the integration of Automatic Identification System (AIS) data and advancements in multimodal and temporal analysis. Unfortunately, due to the lack of freely available high-quality AIS data, this avenue could not be explored in depth. However, securing funding for an AIS data service would open up significant possibilities. AIS data could be directly incorporated into the model as an additional input, enabling regression outputs to estimate the probability of illicit activity for detected ships. Furthermore, incorporating ship-specific features such as size, heading, and speed into the detection model could substantially improve its performance and reliability.

Integration of SAR and EO data offers another promising direction. A multimodal solution that combines Synthetic Aperture Radar (SAR) and Electro-Optical (EO) imagery could enhance robustness and accuracy by leveraging the strengths of both data types. Such a system would be particularly valuable in scenarios where one data modality is unavailable, as it could rely on the other to maintain functionality. Developing a unified model capable of seamlessly handling SAR and EO data would be a critical step toward achieving this goal.

Lastly, incorporating temporal analysis represents an exciting frontier. Extending the model to process sequences of imagery could enable the detection of patterns and anomalies over time, significantly improving maritime domain awareness. This approach would allow the system to predict future ship movements and activities, offering proactive insights into potential threats or illegal operations. These advancements would not only enhance the current capabilities but also broaden the scope of applications for maritime surveillance and security.

VIII. REFERENCES

[1] G. Pili and A. Armenzoni, "Fighting Shadows and Ghosts – A Conceptual Framework for Identification Deception Tactics (IDTs) and MARINE Approach," 2024 IEEE 49th Conference on Local Computer Networks (LCN), Normandy, France, 2024, pp. 1-7, doi: 10.1109/LCN60385.2024.10639801.

[2] Yasir M, Niang AJ, Hossain MS, Islam QU, Yang Q, Yin Y. Ranking Ship Detection Methods Using SAR Images Based on Machine Learning and Artificial Intelligence. Journal of Marine Science and Engineering. 2023; 11(10):1916.

[3] Heiselberg, Peder, Kristian Sørensen, and Henning Heiselberg. "Ship Velocity Estimation in SAR Images Using Multitask Deep Learning." Remote Sensing of Environment, vol. 288, 2023, 113492

[4] Paolo, Fernando S., et al. "xView3-SAR: Detecting Dark Fishing Activity Using Synthetic Aperture Radar Imagery." arXiv, 5 Nov. 2022, https://arxiv.org/pdf/2206.00897.