Ship Port Detection with Coarse and Fine Data Fusion using Edge Boards

S Vasavi, M Sai Pravallika, Y Harini Reddy, D Revanth
[1,2,3,4]Department of Artificial Intelligence and Data Science, V R Siddhartha Engineering College
[1]vasavi_movva@vrsiddhartha.ac.in, [2]pravallika1249@gmail.com, [3]harinireddy123123@gmail.com, [4]revanthdivi3107@gmail.com

Abstract
Accurate detection of ships and ports in satellite imagery remains a significant challenge in maritime security and surveillance due to the complex environmental conditions such as cloud cover, varying image resolutions, and dynamic maritime scenes. To address these challenges, this study proposes a novel method using coarse and fine data fusion integrated with Faster R-CNN (FRCNN), a state-of-the-art real-time object detection framework. By combining multi-scale data, the proposed approach enhances detection precision and robustness, ensuring reliable performance in complex scenarios . A large, annotated dataset of satellite imagery is employed for model training with specialized cloud detection and removal techniques further refining background segmentation and improving detection performance. The solution is optimized for deployment on edge devices, specifically the Jetson Nano, to meet the demands of real-time monitoring in resource-constrained maritime environments. The Jetson Nano, recognized for its compact design and computational efficiency, enables real-time image processing directly at the edge thereby reducing dependence on centralized computing resources. This edge-based architecture ensures both scalability and operational efficiency and is a reliable tool for enhancing maritime security by integrating advanced object detection, data fusion, and cloud mitigation, demonstrating significant potential for real-world surveillance applications.

Keywords. Ship detection, Port detection, Satellite imagery, Coarse and fine data fusion, Faster R-CNN (FRCNN), Cloud detection and removal, Edge computing

Introduction
With the increasing traffic in the world's seas, it has become extremely important to monitor and secure the port regions for commercial as well as strategic reasons. Ports are a very crucial node in international trade and have a high volume of cargo and vessel movements. Hence, it is vital to monitor the area to not only improve operational efficiency but also detect any possible security threats. Traditional monitoring techniques cannot easily scale up with an excellent real-time capability in which such technology has revolutionized a lot. Recently, the source of satellite data was based on multispectral imagery through Sentinel-2A and ResourceSAT-2A, which will show the possibility of improving quality and accuracy in surveillance using a remote sensing-based perspective. Sentinel-2A will capture high-resolution RGB bands for detailed visual information with large-scale monitoring applications. However, ResourceSAT-2A includes other spectral bands, such as NIR and SWIR, that are much richer in spectral information and can be used to differentiate between objects with very similar visual features. This paper presents a Ship Port Detection System based on coarse and fine data fusion with advanced object detection techniques for real-time port monitoring. The methodology employs data from Sentinel-2A and ResourceSAT-2A. Contrast stretching is applied on Sentinel-2A RGB imagery to add NIR and SWIR bands of ResourceSAT-2A to increase spectral differentiation. The deep learning model, Faster R-CNN, is utilized to identify ships and port infrastructure with accuracy in satellite images after preprocessing. With this network architecture of Faster R-CNN, the ships and other features in the port can be effectively located in an accurate manner.

RELATED WORKS

To locate vessels in satellite imagery, [1] evaluates the Faster R-CNN framework: this integrates an RPN with a CNN for robust object detection performance. This significantly improves processing speed through the sharing of computational resources between proposal generation and classification, achieving an mAP of 73.2% on the PASCAL VOC 2007 dataset. With that in mind, the computational cost and memory requirements are rather high, which could reduce its performance while trying to detect small objects or in backgrounds cluttered heavily. The paper focuses the selection of appropriate algorithms for real-time object detection.

Loran et al. (2022) [2] mitigate challenges such as radar noise and low resolution by employing the Faster R-CNN framework to identify ships in range-compressed airborne radar images for ship recognition using airborne radar data, which are often much more complicated to interpret than visual images, hence requiring a model uniquely engineered for such data. With a low false alarm rate of 6.7% and an impressive 91.5% accuracy rate in ship detection, the study shows promise over similar techniques that are tied to traditional radar system. However, this technique suffers from heavy noise interference, and still, it's not apt for real-time application scenarios.

To mitigate the above concerns pertaining to remote sensing images, Rui et al. [3] proposed a Density-aware Cloud Removal technique using a Global-Local Fusion Transformer, named DCR-GLFT. The previous techniques usually complicated cloud removal through reliance on cloud masks or Synthetic Aperture Radar (SAR) data and require large datasets. The DCR-GLFT technique, which uses density information to guide the removal with cloud density estimation, achieves state-of-the-art PSNR = 28.93 and SSIM = 0.84 on the SEN12MS-CR dataset, so this is a major breakthrough for single-image cloud removal regardless of SAR data.

Belgiu et al. [4] report on spatiotemporal image fusion strategies in remote sensing with an emphasis on the combination of sources of satellite imagery, such as radar and optical, respectively. The paper shows how such methods improve spatial resolution and fill in gaps of the temporal coverage so that monitoring capabilities are improved considerably. A few authors have reported increased monitoring and detection accuracy using optical and radar pictures as a combined source, with accuracy gains up to 15% compared to single-source imagery.

PROPOSED METHODOLOGY

8.1. Data Acquisition and Preprocessing

1. *Satellite Data Sources*: For this study, multispectral satellite data was obtained from the following sources
   a. *Coarse Data:* Sentinel-2A was used that provides high-resolution images in various spectral bands such as Blue, Green, and Red bands to identify major features.
   b. *Fine Data:* ResourceSAT-2A, which carries more bands of spectral resolution like the Near Infrared (NIR) and Short-Wave Infrared (SWIR), was used to scan finer spectral details and thereby increasing the spatial resolution in the composite.
2. *Data Import and Georeferencing:* A spectral band for Sentinel-2A was imported separately in.jp2 format, whereas for ResourceSAT-2A, it was in the .tif format. Loading of the bands was done using rasterio, thereby keeping intact the georeferencing information that was essential for spatial alignment between the datasets. This means that pixels in corresponding positions across bands represent the same geographic area.

1.2 Coarse Data Enhancement

To make the coarse data clearer and with more contrast, three spectral bands from the satellite images (Blue, Green, and Red) were loaded as separate bands of grayscale using rasterio. The bands were then stacked in RGB order to create a color composite image. Afterward, normalization(1) [9] was performed in order to scale pixel values between 0 and 1, maintaining consistency between bands and optimizing the data for enhancement.

$$Normalized\ Image = (Image - min\ (Image))/(max\ (Image) - min\ (Image)) \tag{1}$$

To improve the quality of the image as well as visually, we did a contrast stretch using the 2nd and 98th percentiles of the pixel values for each channel independently. This has the side effects of removing outliers yet emphasizes subtle features within channels. The enhanced image has then been saved for further analysis to visualize the improvement in its details. Figure.1 has presented the side-by-side comparison of the original and the enhanced images with the improvement produced from this enhancement process

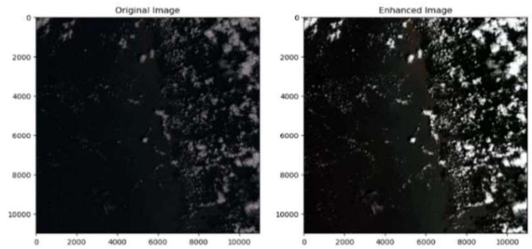

Figure 1. Enhanced Image of Coarse Data

1.3 Fine Data Enhancement

Fine data was also processed with a more detailed approach and more spectral bands. Bands like NIR and SWIR, which provide much more detail for vegetation and water bodies, were used in addition to RGB bands. Multi-band stacking has allowed a richer representation of the scene. The bands were scaled to the same magnitude and principal component analysis used for the enhancement of finer details and removal of redundancy of spectral information.

First three principal components were considered to represent the fine data into RGB, thereby maximizing the spatial and spectral contrast. It is a very helpful transformation in the identification of subliminal features such as health of vegetation, moisture levels in soil, and city details in the fine satellite resolution.

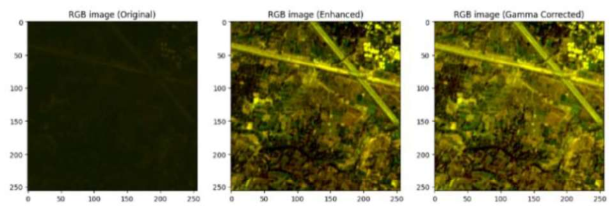

Figure 2. Enhanced Image of Fine Data

1.4 Resampling and Alignment Of Coarse and Fine Data

To verify the spatial alignment between the Coase and Fine datasets, they are introduced to resampling process where the Sentinel-2A bands were resampled to match the resolution of ResourceSAT-2A. This is performed using bilinear interpolation. This helped in reducing data loss and additionally, reprojection was applied where necessary to ensure that the pixels in the same positions across all the bands represented same geographical locations.

1.5 Data Normalization

Following, data normalization was performed to normalize the pixel values across all bands. These are scaled between 0 and 1 to improve consistency across the all bands. This step ensured that the both datasets are of comparable scale and reduced the differences in radiometric resolution.

1.6 Composite Generation

In the next step the normalized coarse and fine data were combined into a single composite image. Weighted averaging method was used where it enhanced both the spatial and spectral features within the composite. These resulting images are served as the input for the further object detection.

1.7 Cloud Mask Generation

Having both coarse and fine data enhanced, a cloud mask [10] was derived that isolates cloud-like regions. That is the necessary step for applications needing cloud-free data. The enhanced image was loaded as a.tif file and was made into grayscale by selecting the red channel, which captures most high-intensity regions to be represented as clouds.

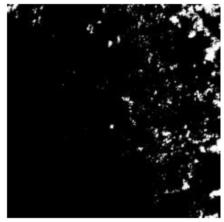

Figure 3. Cloud Masked Data

A simple thresholding technique was used to identify clouds. In this case, all pixel intensities above a predefined threshold of 200 were classified as clouds and set to 255 (white), and all other pixels were classified as non-cloud and set to 0 (black). The resultant binary mask was saved in the .tif format, meaning it would retain all the spatial metadata from the original image and therefore be amenable to geospatial analysis. Figure.3 is the generated cloud mask.

$$Cloud\ Mask = f(x) = \begin{cases} 1 & if\ Pixel \geq Threshold \\ 0 & Otherwise \end{cases} \tag{2}$$

2. DETECTION FRAMEWORK

This section explains the architecture of the enhanced Faster R-CNN model as shown in Figure.4, how each layer assisted in the precise detection of ships and ports in the satellite imagery.

Input Layer: The input is the enhanced, cloud-free image obtained from various preprocessing steps including cloud masking. This layer ensures that the enhanced model focuses only on the essential details required for a precise detection.

Backbone CNN(VGG Network): The backbone VGG Network helped in extracting the necessary features from the input image, ie both the low-level(edges, textures) and high-level features. This feature extraction helped in ships from each other using their unique shapes, different layouts of ports even in the complex backgrounds.

Region Proposals (RPN): This layer created the potential regions of interest within the image. These are the areas that are likely to contain ship or port features. This essential step ensured that the model only focusses on these high-probability areas consequently improving the detection accuracy and its computational efficiency.

ROI Pooling Layer: This ROI Pooling layer takes the generated regions of interest and normalizes them to a uniform size. This step ensured that all the images such as large ships and different port areas are processed evenly and resulted in consistent detection across various scales.

FC Layer: After pooling, the model learns about the generated features and assigns each region of interest a category such as ship or port, localize each object through bounding box regression.

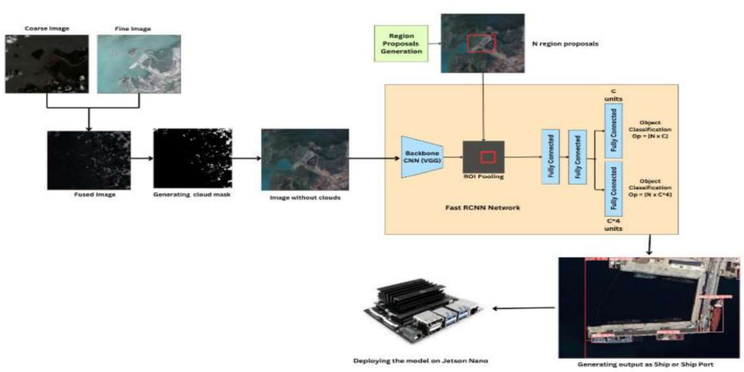

Figure 4. Faster R-CNN model Architecture

3. RESULTS AND ANALYSIS

The performance metrics for the enhanced model include precision, recall and mAP at IoU. These results provided an analysis on how the model performed in various cases. As shown in Table 1, the model achieved an overall recall of 0.73 and a precision of 0.83 on all instances.

Metrics	Ships	Ports	Average
Precision	0.85	0.82	0.835
Recall	0.76	0.71	0.735
F1 Score	0.80	0.76	0.78
Mean IOU	0.68	0.65	0.665
Total Predictions	30	25	55

This showed that there was a balanced performance and even successfully minimized the false positive rates. For the ship class, the precision and recall rates strongly show that the model captured various instances although it missed a few true positives at some points. Additionally, the mean IoU values for both classes indicates that the model achieved a reasonable accuracy in detecting the objects along with bounding boxes that closely matched with the true positions of ships and ports.

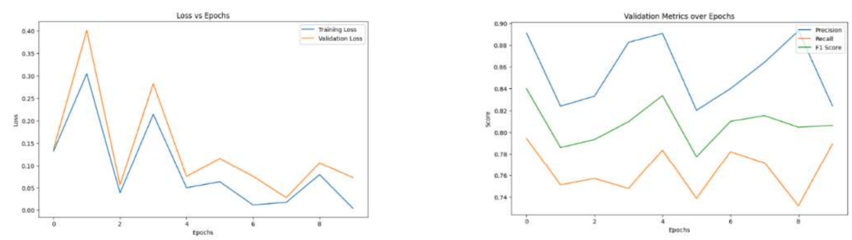

Figure 5. Loss vs Epochs

Additionally, the validation metric graphs provided the models performance over the training epochs. As depicted in Figure.5 although recall displayed few instabilities over a range, precision remained constantly high indicating the models accuracy in detecting ships and ports and the loss curves declining over the epochs displayed the models effective learning and generalization. During the initial stages the model showed low values but in later epochs there was a strong convergence.

Figure 7. Results of Ship Port Detection

Overall all these rigorous steps helped in bringing out a detailed analysis of the model and its effectiveness across various environments, pointing out its potential in real-time scenarios and for various practical applications in maritime monitoring and management.

4. IMPLEMENTATION ON JETSON NANO

Compact computer modules like the NVIDIA Jetson Nano are made for edge AI applications, especially in fields like robotics, drones, and Internet of Things devices. It is the perfect environment for developers to test out AI applications because it uses GPU architecture to effectively implement deep learning algorithms without depending on cloud services. Installing required libraries, downloading pre-trained weights, and setting up an appropriate environment are all part of configuring Faster RCNN for object identification on the Jetson Nano. Faster RCNN overlays bounding boxes on images and recognizes objects during inference. Light weight model like TensorRT is used to boost performance. This model incorporates methods like quantization to increase processing speed without sacrificing accuracy.

Table 5: Inference in different scenarios

Power Consumption	Inference Time
5W	36298.4ms
MaxN	16434.2ms

Based on power consumption choices, performance evaluations show notable variations in inference times. For example, the system may operate 2.2 times faster (16,434.2 ms) in "MaxN" mode than in low power "5W" mode (36,298.4 ms), suggesting that inference speed is directly improved by higher power available. The system operates 2.2 times faster in "MaxN" mode than in 5W mode, according to the analysis of the data from Table 5. This discrepancy is probably caused by the system's ability to pull power in "MaxN" mode, which improves performance.

5. CONCLUSION AND FUTURE WORK

This study addresses the difficulties presented by complicated maritime settings by proving the efficacy of the suggested method, which makes use of an improved Faster R-CNN model for ship and port detection in satellite data. Faster R-CNN's incorporation of coarse and fine data fusion greatly enhances detection robustness and precision, making it appropriate for practical uses in marine security and surveillance. There is potential for improvement even though the Jetson Nano's CPU's current implementation offers respectable inference speeds.

Future research will concentrate on optimizing the suggested model's performance using a variety of Jetson Nano-specific optimization strategies. While investigating smaller versions of the Faster R-CNN model may result in better inference times, tracking GPU consumption will provide insights for future improvements.

6. REFERENCES

[1] S. Ren, "Faster r-cnn: Towards real-time object detection with region proposal networks.," arXiv preprint, no. arXiv:1506.01497, 2015.

[2] "Ship detection based on faster R-CNN using range-compressed airborne radar data.," IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1-5, 2022.

[3] R. Quan, S. He, T. Li, G. Wang, N. Ruan, M. Lin, Y. Yang and H. Shen, "Density-aware Cloud Removal of Remote Sensing Imagery Using a Global-Local Fusion Transformer," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[4] M. Belgiu and A. Stein, "Spatiotemporal image fusion in remote sensing," Remote sensing, vol. 11, no. 7, p. 818, 2019.

[5] Y. Chen, M. Chen, W. He, J. Zeng, M. Huang and Y. Zheng, "Thick cloud removal in multitemporal remote sensing images via low-rank regularized self-supervised network," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[6] R. Dong, L. Zhang, W. Li, S. Yuan, GanL, J. Zheng, H. Fu, L. Mou and X. Zhu, "An adaptive image fusion method for sentinel-2 images and high-resolution images with long-time intervals," International Journal of Applied Earth Observation and Geoinformation, vol. 121:103381, 2023.

[7] Y. Wu, W. Ma, M. Gong, Z. Bai, W. Zhao, Q. Guo, X. Chen and Q. Miao, "A coarse-to-fine network for ship detection in optical remote sensing images," Remote Sensing, vol. 12(2):246, 2020.

[8] S. He, H. Zou, Y. Wang, R. Li, F. Cheng, X. Cao and M. Li, "Enhancing mid–low-resolution ship detection with high-resolution feature distillation," IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2021.

[9] M. Hall-Beyer, "GLCM texture: a tutorial," National Council on Geographic Information and Analysis Remote Sensing Core Curriculum, vol. 3, no. 1, p. 75, 2000.

[10] S. Ackerman, K. Strabala, W. Menzel, R. Frey, C. Moeller and L. Gumley, "Discriminating clear sky from clouds with MODIS," Journal of Geophysical Research: Atmospheres, vol. 103, no. D24, pp. 32141-32157, 1998.