iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.63005

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Tiny Ship Detection in Remote Sensing Images

Anurag Nikam[1], Mrs Poonam Jadhav[2], Anuj Karade[3], Daanish Shaikh[4], Ajay Singh[5]
[1, 3, 4, 5]*UG student,* [2]*Professor, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India*

Abstract: ***Due to the difficulties of complicated settings, it is challenging to detect rotating ships in optical remote sensing photographs. Advanced rotational ship detectors currently in use are algorithms which are anchor-based that call for a large number of preconfigured anchors. However, using anchors has these serious issues: Ad hoc heuristics are used to develop anchor characteristics, e.g. size and aspect ratio. Positive samples are only possible for a very tiny percentage of anchors that closely overlap with the bounding boxes of ships.This causes a large disparity in the proportion of positive and negative samples. Poor anchor design will therefore significantly affect the accuracy of detection. This research suggests a unique framework, to solve the aforementioned issues by identifying ships as keypoints in optical remote sensing images. Ship targets in SKNet are modelled using an object's width, height, rotation angle, and centre keypoint. Given this, we develop two distinct modules: orthogonal pooling and softrotate-nonmaximum suppression (NMS). The former seeks to improve the morphological size and centre keypoint predictability, while the latter effectively removes redundant rotated ship detection results.***
Keywords: Object detection, ship detection, convolutional neural network, YOLOv8, image processing, deep learning.

I. INTRODUCTION

SHIP detection in optical remote sensing photos is critical for traffic surveillance and other application like maritime security and prevention from illicit fishermen. Due to the recent rapid evolution of RS technologies, the availability of multiple high-resolution RS photographs [2] has greatly assisted ship identification [3]–[6]. Additional challenges include the complex sceneries, odd shooting angles, and different object sizes in a single shot.

Ship pose estimation, marine traffic control, and other applications depend heavily on the consensuses that may be used to detect the ship targets. Even though the aforementioned detectors are effective in most situations, there are some specific instances when axis-aligned detection is unable to accurately determine the item's true position. Depending on the geographic location, general detection boxes usually contain a large number of background pixels and even regions that are intended to belong to the close objects.

This uncertainty can be minimised by carefully fitting the boundaries of ship targets with rotational boxes. Pose estimation and trajectory prediction are therefore made feasible. Even Nevertheless, rotating an object makes it rather difficult to identify it because it necessitates anticipating not just its position and size but also its rotation angle. These methods involve stretching the anchor. While current approaches have made some progress in rotational ship recognition in optical remote sensing images, these detectors' performance is still too low for practical usage.

II. LITERATURE REVIEW

Hao Li et al. [1] improved the v3 small network of YOLO to facilitate the localization and detection of six different types of ships. To simplify model training, they introduced the first six new anchors specifically tailored to match the designs of the ships. Aim of this adjustment was to optimize the performance and accuracy of model in detecting the specified ship types.

Jiang Kun et al. [2] introduced a noteworthy enhancement to the Yolov4-tiny algorithm, effectively boosting its accuracy. They suggest adding an attention mechanism unit, which significantly improves the model's capacity to identify small targets. This is a major improvement in object detection, especially in situations when accurate small object identification is essential.

Jiang Feng et al. [3] shown how the RpiFire technology has a high recall rate and exceptional accuracy. Positive results from picture and video testing, as well as simulated fire trials carried out in an interior space free of shiners, were reported. This demonstrates the system's efficacy in detecting fires and its potential for practical uses.

Ting Yang et al. [4] introduced YOLOv3, an improved version created especially for automatic ship detection. Their work focuses on using optical ship datasets and Synthetic Aperture Radar (SAR) to improve the model's performance. With this updated version, the goal is to improve ship recognition accuracy and robustness, particularly in situations when many image modalities are involved.

Zhenyu Cui et al. [5] presented SKNet, a cutting-edge framework designed to identify ships in optical remote sensing pictures. Using a unique technique, this framework depicts every ship as a center focal point with matching morphological sizes.

©IJRASET: All Rights are Reserved

SKNet hopes to enhance the precision and efficacy of ship detection in optical remote sensing pictures by employing this modelling technique, providing a workable solution for a range of maritime surveillance and monitoring applications.

Shiqi Chen et al. [6] recommended that the tiny and effective Tiny YOLO-Lite object detector be used for fast and accurate detection of ships in SAR (synthetic aperture radar) images. The suitability of Tiny YOLO-Lite for effectively recognizing ships in SAR data is highlighted by this recommendation, highlighting its potential for real-time maritime surveillance and monitoring applications.

Haitao Lang et al. [7] revealed a new technique for identifying ships and is limited to High Resolution single-channel SAR (synthetic aperture radar) images. In order to increase accuracy and efficiency in maritime surveillance applications, this work aims to overcome the challenges related to ship detection in HR SAR images.

Alifu Kuerban et al. [8] developed a novel multi-vision small object recognition framework with the goal of quickly and precisely identifying different kinds of vehicles, such as automobiles, ships, and aircraft, from remote sensing photographic data. This advanced system is intended to improve item identification jobs' accuracy and efficiency, especially in the context of remote sensing applications where small object detection presents substantial obstacles. This detector is positioned to help advance object recognition capabilities in remote sensing imagery by combining sophisticated algorithms and techniques. Numerous uses will be made possible by this, including monitoring of transportation and the environment.

Z.A. Khan et al. [9] provided a real-time system that combines the PCA (Principal Component Analysis), RT (Radon transform), and LDA (Linear Discriminant Analysis) to identify human activities. This system uses ANN (Artificial Neural Networks) to distinguish between various human activities.

Agadeesh B et al. [10] have put forth a process that uses information from optical flow analysis to create a binary image. Afterwards, feature vectors are extracted from these binary pictures using the HOG (Histogram of Oriented Gradient) descriptor. This technique may be used to build a robust model that effectively classifies data based on the features that are gathered, which makes it a valuable tool for a number of applications such as object tracking, motion analysis, and picture identification.

Aishwarya Budhkar et al. [11] Examine action recognition in videos, especially using the KTH dataset. They investigate four feature descriptor and classifier combinations: 3D SIFT and HOG descriptors combined with SVM and KNN classifiers. A Bag-of-Words model is constructed by extracting features from training video frames and clustering them together. This approach enables a comprehensive evaluation of several classifiers and descriptors, providing insightful information for improving recognition algorithms in action-based videos.

Chengbin Jin et al. [12] are looking into a way to use CNN (Convolutional Neural Networks), a class of deep learning models which is capable of independently learning features from training movies, to predict human actions using temporal imagery. Current state-of-the-art approaches, while demonstrating exceptional accuracy, frequently demand large processing resources. Moreover, a prevalent issue results from many approaches presuming an accurate comprehension of human postures.

Olliffe et al. [13] have backed the basic idea of principal component analysis (PCA), which aims to preserve a sizable amount of the underlying variance while reducing the dimensionality of a dataset with several related variables. By converting the data into a new set of variables known as principle components (PCs), this objective is accomplished. These elements are uncorrelated and ranked so that the majority of the variation seen in the original variables is retained first, with special attention to the first few elements.

E. Mohammadi et al. [14] are investigating the use of an ensemble of support vector machines (SVMs) for Human Activity Recognition (HAR) in order to enhance classification performance by merging various features from various angles. They use Dempster-Shafer fusion and the product rule from algebraic combiners to combine the outputs of various classifiers.

T. B. Hilton et al. [15] highlighted the relevance and ongoing growth of human motion capture in computer vision research, as demonstrated by the volume of over 350 articles in the last few years. They outlined several noteworthy developments in research, such as novel approaches to tracking, pose estimation, automatic initialization, and movement identification. In particular, recent efforts have concentrated on attaining precise tracking and location estimates in scenes that are naturally occurring. Furthermore, a great deal of progress has been achieved in the automatic comprehension of human behavior and behaviors.

III. PROPOSED SYSTEM

1) Preprocessing:

- To begin, compile a dataset of remote sensing photos that includes both ship-free and little ship shots.
- Add annotations to the dataset by drawing bounding boxes around the small ships to give accurate location data.
- The dataset should then be divided into distinct subsets for training, validation, and testing.

- Execute preparation activities for images, such as scaling, normalizing, and using rotation and flipping as data augmentation methods. These actions are essential for improving the training data's robustness and diversity, which will enable more efficient model training.

2) Model Selection:

- Choose the model architecture or a similar deep learning model suitable for object detection tasks.
- You can implement the model from scratch or use pre-trained models for transfer learning, depending on the availability of data and computational resources.

3) Model Training

- Initialize the selected model with random weights or load pre-trained weights if applicable.
- Define loss functions for object detection, such as mean squared error (MSE) or more advanced ones like focal loss. • Implement an optimizer (e.g., Adam, SGD) to update model weights.
- Train the model for multiple epochs while monitoring the loss on the validation dataset to prevent overfitting.
- Utilize techniques like learning rate scheduling and early stopping to improve training stability.

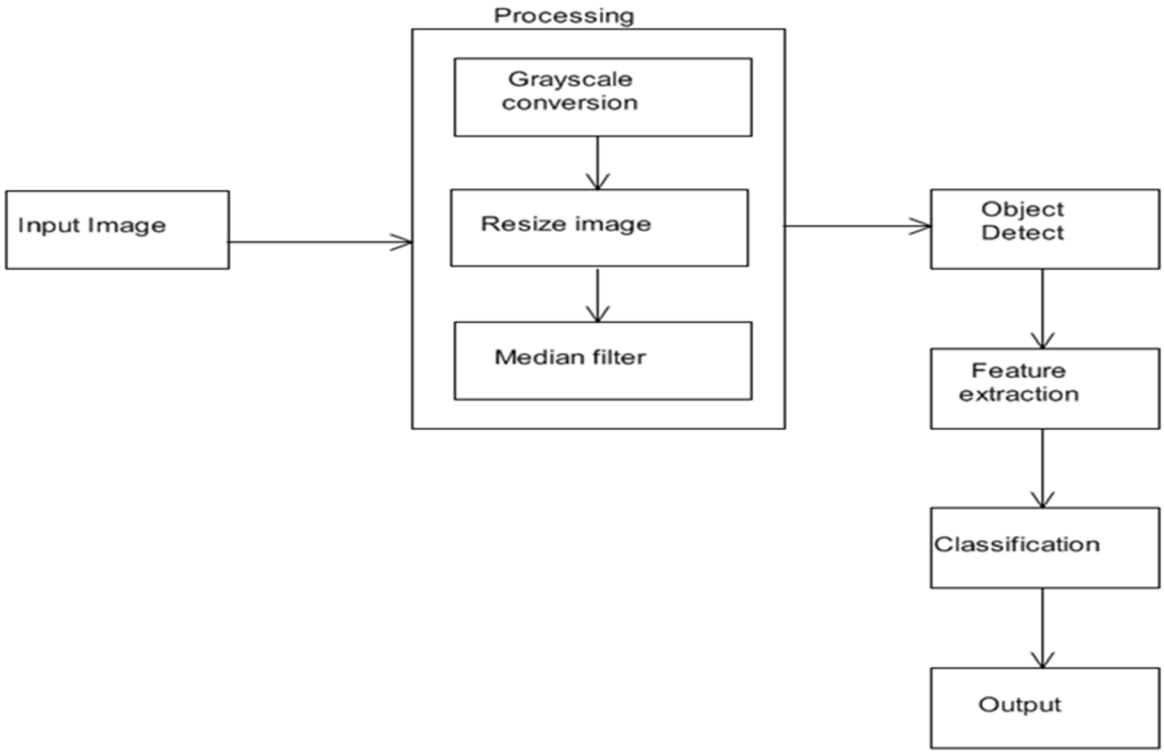


Figure 1. System Architecture

4) Object Detection:

- After training, your model will be capable of detecting tiny ships in images.
- Use non-maximum suppression (NMS) or other post-processing techniques to remove duplicate or overlapping detections.

IV. ALGORITHMS

1) YOLO takes an input image by breaking it down into a grid.

2) Each cell in the grid-like representation of the image is tasked with forecasting the locations of objects.
3) Predictions regarding object classes are made for each bounding box, estimating the likelihood of an object belonging to a certain class.
4) Utilizing a basic deep convolutional neural network, YOLO processes the image to detect objects. The structure of the CNN model serving as the foundation for YOLO is depicted in the accompanying figure.

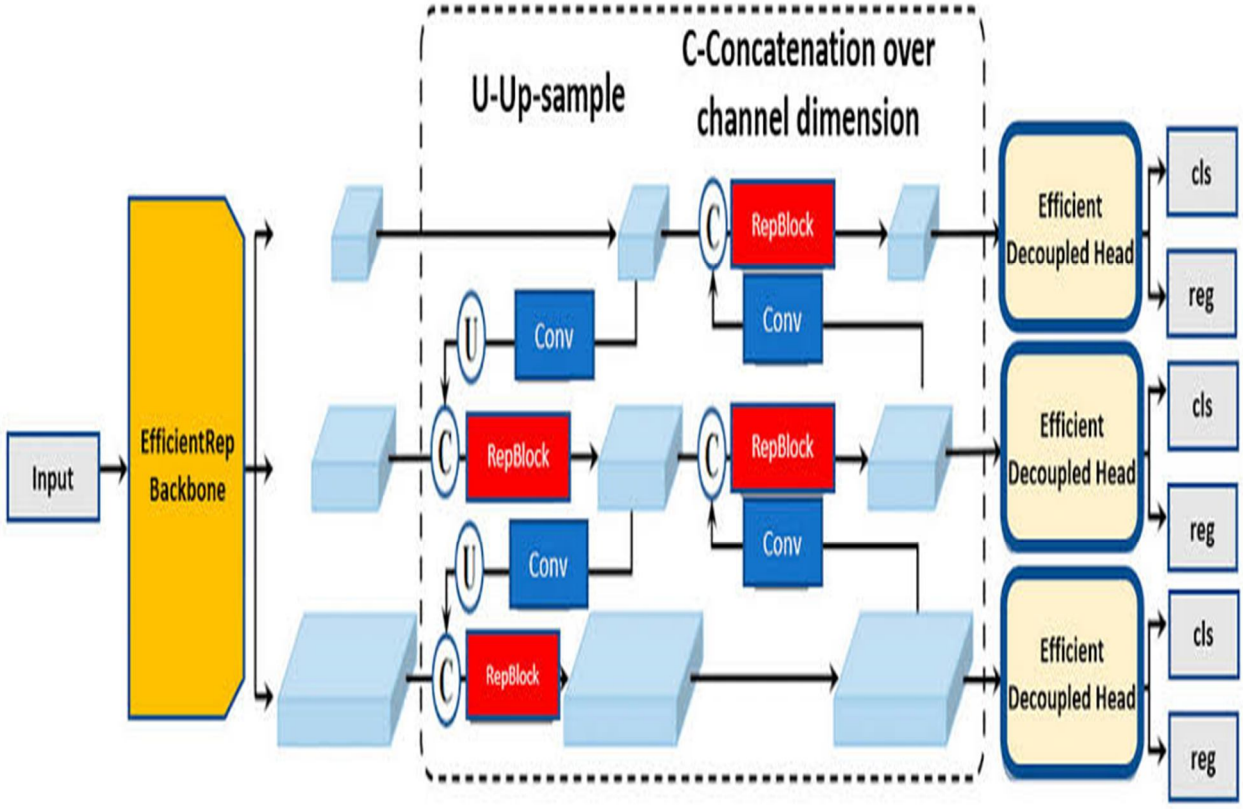
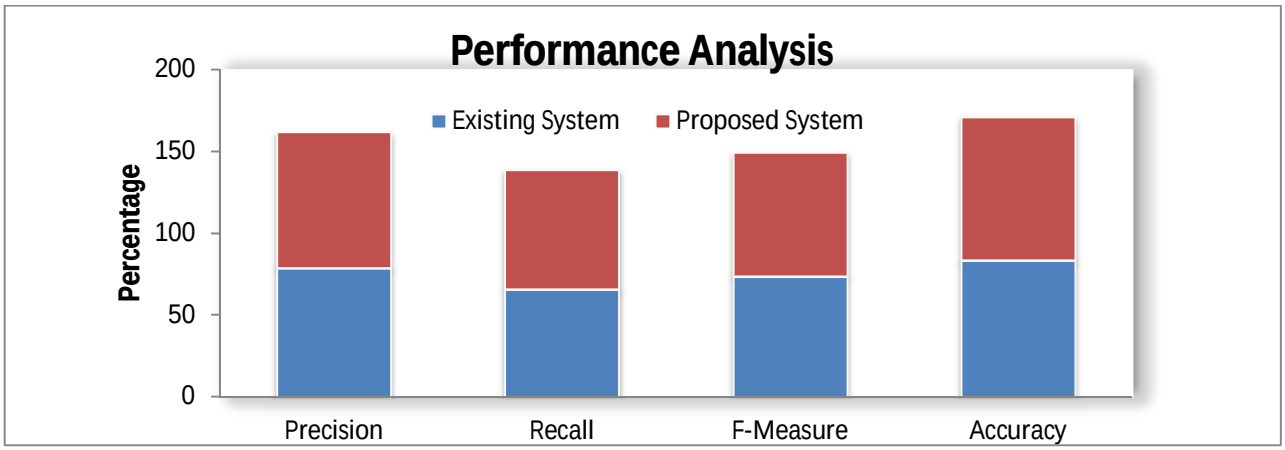


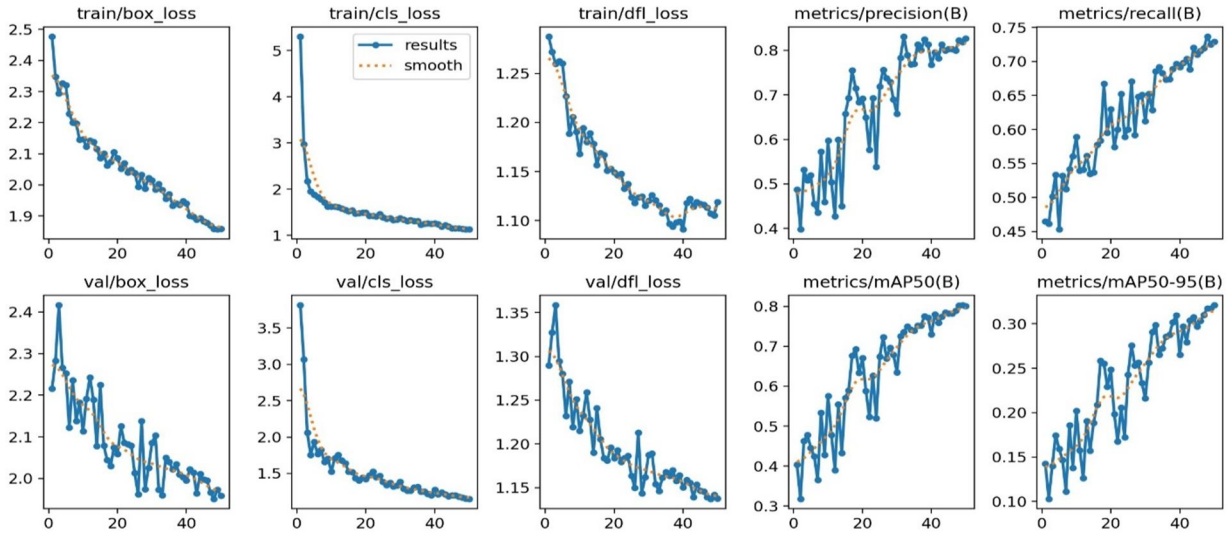
Figure 2. Yolo Working

5) Using ImageNet, the first 20 convolution layers of the model are pre-trained by inserting a fully linked and temporary average pooling layer. Subsequent research has demonstrated that incorporating convolution and connected layers into a pre-trained network enhances its performance; hence, this pre-trained model is transformed to conduct detection. The last fully connected layer of YOLO predicts bounding box coordinates as well as class probabilities.
6) Yolo partitions an input image into a S × S grid. An object's centre falls into a grid cell, and that grid cell is in charge of detecting it. B bounding boxes and confidence scores for those boxes are predicted for each grid cell. The model's level of confidence that the box includes an object and the accuracy of the anticipated box are both indicated by these confidence ratings.
7) For each grid cell, Yolo predicts numerous bounding boxes. We only want one bounding box predictor to be in charge of each object during training. Depending on whose prediction has the highest current IOU with the ground truth, YOLO designates one predictor as "responsible" for making an object prediction. As a result, the bounding box predictors become specialised. By enhancing its ability to forecast specific object sizes, aspect ratios, or classes, each predictor raises the recall score as a whole.
8) The YOLO models employ NMS (non-maximum suppression) as a crucial method. NMS is a post-processing stage that raises object detection's precision and effectiveness. It is typical practice in object detection to generate numerous bounding boxes for a single object in an image. The bounding boxes indicate the same object, even though they may overlap or be positioned differently. NMS is used to extract a single bounding box for each object in the image and to find and eliminate unnecessary or inaccurate bounding boxes.
9) Now explore the enhancements introduced by the later iterations of YOLO to its original model.

V. RESULT ANALYSIS



	Existing System	Proposed System
Precision	78.62	82.70
Recall	65.43	72.90
F-Measure	73.10	74.31
Accuracy	82.29	87.26



VI. CONCLUSION

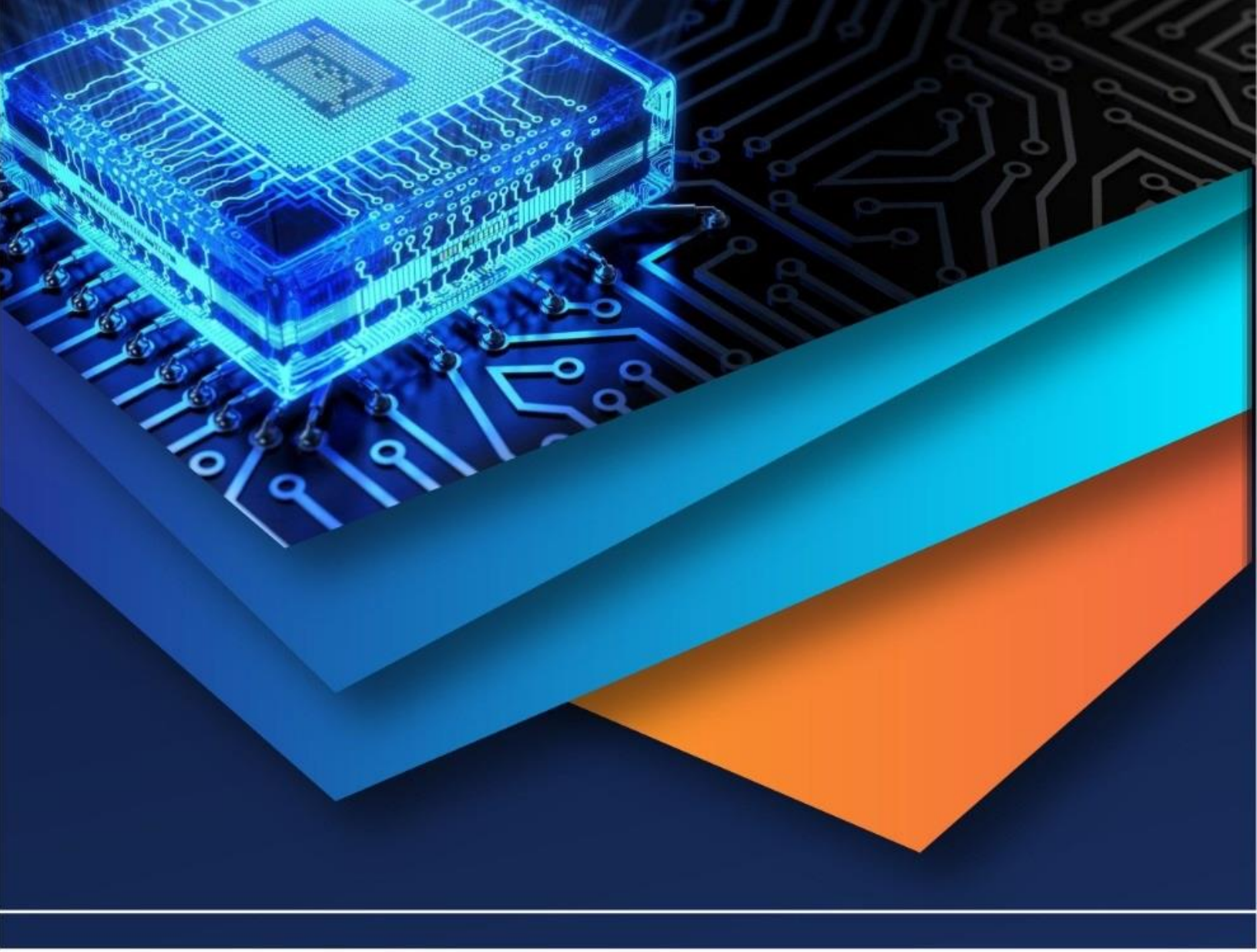
The current project intends to propose an extremely accurate ship detecting system based on YOLOv8. It is anticipated that the algorithm would exhibit exceptional flexibility in responding to diverse environmental conditions and will be able to detect ships in real time when necessary. The research project requires the development of thorough performance indicators, documentation, and a framework for continuous enhancement in order to improve maritime surveillance capabilities and meet stakeholder objectives. The study effort hopes to improve marine surveillance capabilities by accomplishing these goals, which could have a big impact on a lot of academic and business environments.

REFERENCES

[1] Hao Li , Lianbing Deng, Cheng Yang, Jianbo Liu, and Zhaoquan Gu, “Enhanced YOLO v3 Tiny Network for Real-Time Ship Detection From Visual Image”, January 25, 2021.
[2] Jiang Kun, Coa Yan, “SAR Image Ship Detection Based on Deep Learning”, 2020 ICCIC.

[3] Jiang Feng,Yang Feng, Wu Benxiang, “Design and experimental research of video detection system for ship fire”, June 19,2020.
[4] Ting Yang, Xiaohua Tong, Senior Member, “Multi-Scale Ship Detection from SAR and Optical Imagery via A More Accurate YOLOv3”, 2021.
[5] Zhenyu Cui, Jiaxu Leng, “SKNet: Detecting Rotated Ships as Keypoints in Optical Remote Sensing Images”, 2021.
[6] Shiqi Chen, Ronghui Zhan, Wei Wang and Jun Zhang, “Learning Slimming SAR Ship Object Detector through Network Pruning and Knowledge Distillation”, 2020.
[7] Haitao Lang, Yuyang Xi, “Ship Detection in High-Resolution SAR Images by Clustering Spatially Enhanced Pixel Descriptor”, 2019.
[8] Alifu Kuerban, Yuchun Yang, Zitong Huang, Binghui Liu, and Jie Gao, “MultiVision Network for Accurate and Real-Time Small Object Detection in Optical Remote Sensing Images”, 2021.
[9] Z.A. Khan, W. Sohn, “Real Time Human Activity Recognition System based on Radon Transform”, IJCA Special Issue on Artificial Intelligence Techniques, 2011.
[10] Agadeesh B, Chandrashekar M Patil, “Video Based Action Detection and Recognition Human using Optical Flow and SVM Classifier”, IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016.
[11] Aishwarya Budhkar, Nikita Patil, “Video-Based Human Action Recognition: Comparative Analysis of Feature Descriptors and Classifiers”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 6, June 2017.
[12] ChengbinJin, ShengzheLi, Trung Dung Do, Hakil Kim, “Real-Time Hu- man Action Recognition Using CNN Over Temporal Images for Static Video Surveillance Cameras”, Information and Communication Engineering, Inha University,Incheon,Korea,2015.
[13] Olliffe, I.T., “Principal component analysis”, Springer Series in Statistics,2nd ed., Springer, 2002.
[14] E. Mohammadi, Q.M. Jonathan Wu, M. Saif, “Human Activity Recognition Using an Ensemble of Support Vector Machines”, IEEE International Conference on High Performance Computing Simu- lation (HPCS) , July 2016.
[15] T. B., Hilton, A., and Kruger, V, “A survey of advances in vision-based human motion capture and analysis”, Computer Vision and Image Understanding, vol. 104, pp. 90-126, 2006.
[16] T. Misaridis and J. A. Jensen, “Use of modulated excitation signals in medical ultrasound. Part III: high frame rate imaging”, IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 52, no. 2, pp. 208-219, Feb. 2005.
[17] R. J. Zemp, A. Sampaleanu, and T. Harrison, “S-sequence encoded synthetic aperture B-scan ultrasound imaging”, IEEE International Ultrasonics Symposium (IUS), Jul 2013, pp. 593-595.
[18] N. Bottenus, “Recovery of the complete data set from focused transmit beams”, IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 65, no. 1, pp. 30-38, Jan 2018.
[19] N. Bottenus, “Recovery of the complete data set from ultrasound sequences with arbitrary transmit delays”, J. Acoust. Soc. Am., vol. 31, no. 1, p. 020001, Dec. 2017.
[20] J. Liu, Q. He, and J. Luo, “A compressed sensing strategy for synthetic transmit aperture ultrasound imaging”, IEEE Trans. Med. Imag., vol. 36, no. 4, pp. 878-891, Apr. 2016.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)