SAR Image based Target Detection & Identification using Deep Learning Model YOLOv7

B Nehru[1]*, Vaddi Chandra Sekhar, A Naresh, D Penchalaiah, Abhijit Bhattacharya

Department
DRDL
Email: nehrurgukt8@gmail.com

Abstract

In this paper we have implemented a Real Time ship Target Detection on SAR Imagery Using deep learning model. Ship detection is crucial to neutralize marine targets during combat scenario. SAR images operate in all-weather conditions and all-day and independent of range. SAR imagery-based surveillance technology is particularly well suited for marine surveillance. An Open-source SAR-ship dataset (HRSID) along with custom dataset is used for detection and identification. The existing AI based methods of detection (two stage object detectors) have proposed regions followed by extraction of features. The speed of the process for these methods is slow for real-time applications. Hence one stage detector, YOLOv7 with 415 convolution layers is used for the application which will overcome constraints of the previous methods. Mean Average Precision (mAP) of 78% @ 0.5 IOU is achieved for test HRSID dataset and custom dataset. NVIDIA GPU (RTX A4000) hardware is used for the training. The inference time of 8.2 milliseconds is achieved on NVIDIA RTX A4000 GPU. Hence YOLOv7 is suitable for real time detection of military targets on SAR images.

Keywords: *IOU, Convolution networks, SAR, mAP*

1. Introduction

Radar is used in military applications for remote sensing, target detection, guidance and global positioning [1]. Radar has a limitation of physical dimension of aperture [2]. SAR (Synthetic Aperture Radar) is a technique to overcome the physical limitation of radar by using signal processing. SAR is able to illuminate wide area for 24x7 in any weather condition with wide look angles. So, it is very useful for military applications where detection of objects is important for situational awareness.

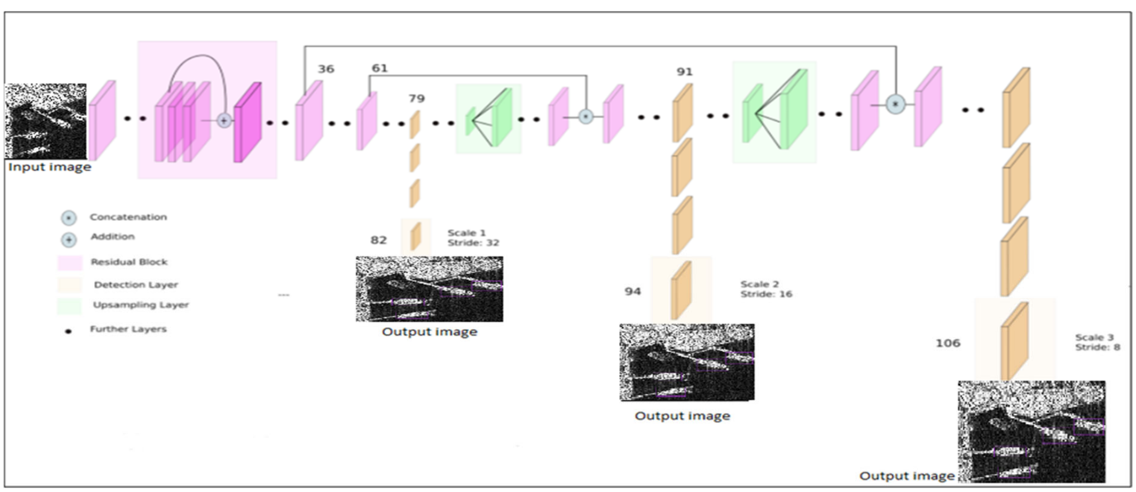

Figure 1: One stage detector of YOLOv7

[1] Corresponding author: E-mail: nehrurgukt8@gmail.com

For detection of objects on SAR imagery, YOLOv7 convolutional network is used. YOLOv7 [3] is single stage detector where detection and identification of objects is done in one pipeline, so it is suitable for real time applications compared to two stage detectors. Architecture of one stage detector is show in Figure 1. YOLOv7 can be trained on custom dataset without pre-trained weights. Aim of this paper is to detect ship target from SAR images. Open source dataset HRSID (High resolution SAR image dataset) along with custom dataset of 800x800 pixels size images is used for training the model. Convolutional neural network (CNN) is deep, feed-forward neural network for identifying objects from visual images. CNN includes convolutional layer, fully-connected layers (Dense layer) and pooling layer. In existing two stages CNN models' regions are proposed and then given for the classifier for detection of object, which makes them slow. YOLOv7 outperforms the older versions of YOLO in speed (prediction & training) and accuracy. The training & inferencing of YOLOv7 is done on NVIDIA RTX A4000 GPU.

2. Literature Survey

A single stage detector model, YOLOv7 is used for detection of ships on SAR images. Wang et. Al proposed yolov7 algorithm. They architecture of YOLOv7, combines the strengths of Scaled-YOLOv4 and EfficientDet-D7 models to achieve high accuracy and fast performance. They have proposed the trainable bag-of-freebies method to enhance the accuracy of object detection. Vo et. al [4] used CeyMo Road Marking dataset for detection and identification road markings. They have experimented and evaluated performance of the different object detection models and algorithms. They could find out that YOLOv7 performed better compared to other methods and it is current best method for effective detection and identification of road markings for their road marking dataset. Liu et. Al [5] investigated to build a robust and fast model for estimation of vehicle distance. Synthesized theoretical insights in the context of odometry and based on proposed conceptual framework YOLOv7 is customized. In this research, KITTI data set is used for distance estimation using their attention model along with YOLOv7, YOLOv7 is used for the first time. Gallo et. al [6] performed deep object detection using real case dataset of crop weeds captured from uav images. The YOLOv7-based object detection model performed better compared to the existing two-stage detectors (RCNN, FRCNN) and single-stage (YOLOv5, YOLOv4) in terms of detection accuracy. YOLOv7 achieved better results compared to YOLOv5 in terms of the mAP score using the LB dataset. L Jin et al [7] proposed a real-time fault diagnosis model of PV modules for integrated energy systems to overcome the complexity of infrared images of PV array based on YOLOv7. They have shown that compared to SSD [8] and Faster-RCNN [9] target object detection algorithms YOLOv7 has a faster image processing speed.

3. Architecture Overview

YOLOv7 (You Only Look Once version 7) is a state-of-the-art object detection algorithm that uses deep learning to detect objects in images and video. The block diagram is shown in figure (2). The architecture is an improvement over its predecessor YOLOv5 [10], and is based on a combination of the Scaled-YOLOv4 [11] and EfficientDet-D7 models.

YOLOv7 has following key features in its architecture:

Backbone network: YOLOv7 uses the CSPDarkNet53 as its backbone network, which is a lighter and faster version of the DarkNet architecture. The CSPDarkNet53 includes cross-stage partial connections, which allow for better feature propagation and optimization.

Neck network: YOLOv7 uses the SPP-PANet as its neck network, which includes a Spatial Pyramid Pooling (SPP) module and a Path Aggregation Network (PAN). The SPP module allows for detection of objects at different scales, while the PANet module aggregates feature maps from different levels to improve object detection accuracy.

Head network: YOLOv7 uses the YOLOv4-Lite as its head network, which includes a series of convolutional layers followed by a detection layer. The detection layer uses anchor boxes and class probabilities to generate object detection predictions.

Training strategy: YOLOv7 uses a combination of supervised and unsupervised learning strategies, including data augmentation, mix up training, and self-training. The model is trained on large-scale datasets such as COCO [12], Pascal VOC [13], and ImageNet [14].

Overall, the YOLOv7 architecture is designed for fast and accurate object detection, with a focus on optimizing the performance of the backbone, neck, and head networks. It achieves state-of-the-art results on object detection benchmarks such as COCO and KITTI, making it a popular choice for various applications such as autonomous driving, surveillance, and robotics.

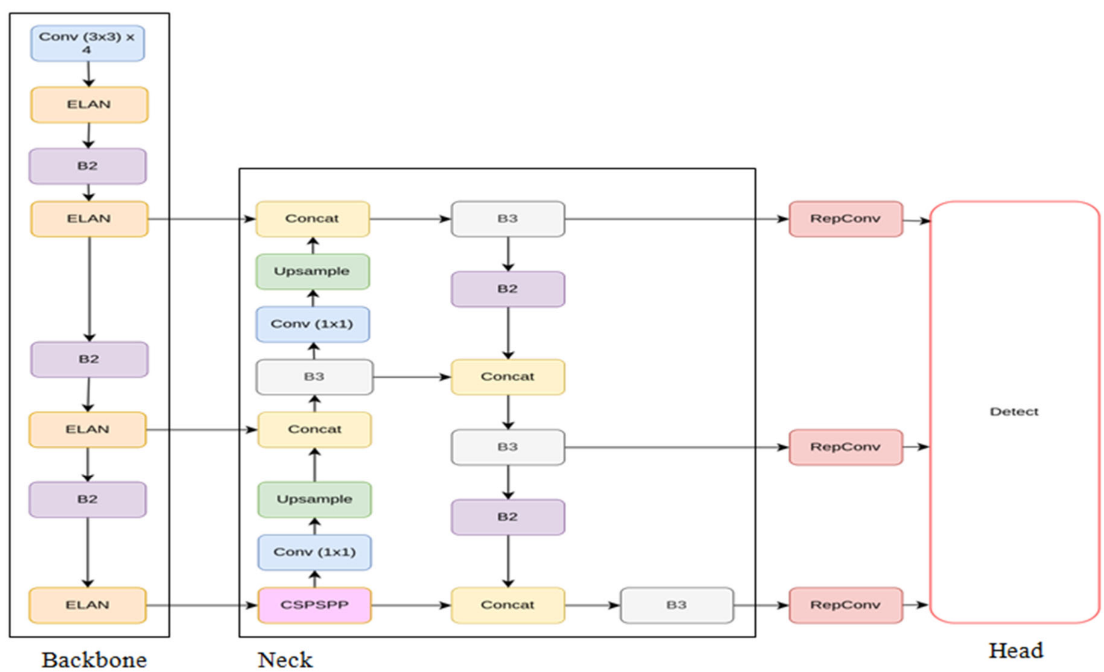

Figure 2: YOLOv7 Architecture

4. Real Time Target Detection for SAR Imagery

Dataset

HRSID open dataset & custom dataset is used for training. During training, 3000 images are used for training, 1647 images are used for validation and 300 images are used for testing. The following are the few images from dataset.

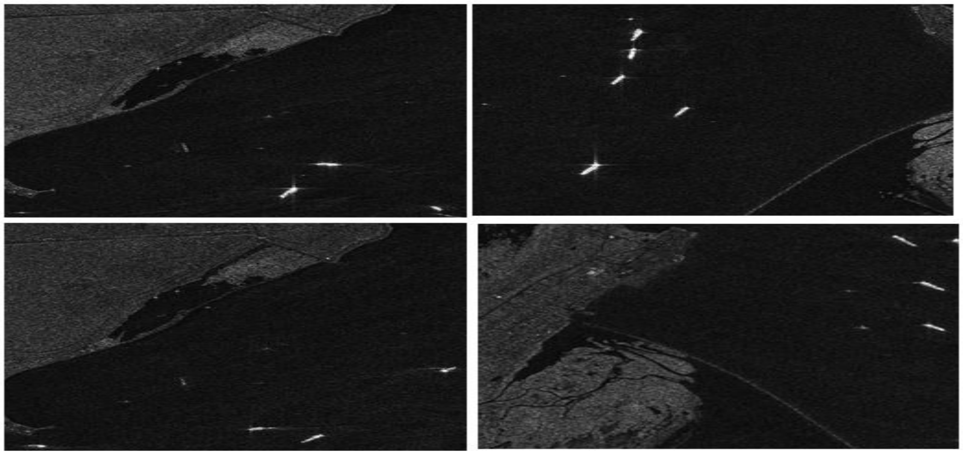

Figure 3: Images from SAR-ship dataset

DATASET ANNOTATION

Dataset is annotated in YOLO Format. In YOLO format center coordinates and width and height which represent bounding box for each object in an image. LabelImg, which is graphical user interface tool is used for annotating SAR images of ship. YOLO format contains, center coordinates, length and width of bounding boxes.

5. EXPERIMENTAL RESULTS

DETECTION PERFORMANCE ANALYSIS OF THE YOLOV7 MODEL

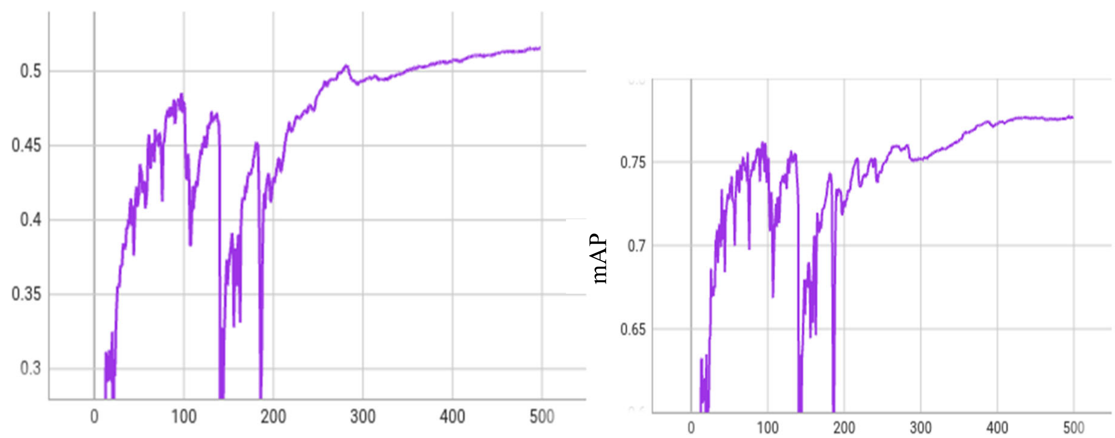

Fig.4 (a) mAP@ IOU 0.5 Epochs Fig.4 (b) mAP@ IOU 0.5-0.95 Epochs

Ship target detection and identification is carried out on NVIDIA RTX A4000 on Ubuntu 20.04 of Linux operating system. NVIDIA GPU has GPU memory 16GB, 6144 Cuda cores. The model performance is measured in terms of mean Average Precision (mAP). The model achieved 0.78 mean average precision (mAP) with validation dataset of size 800 x 800 as depicted in figure 3. An mAP @ IOU 0.5 indicates mean average precision at IOU of 0.5. Model is trained for 500 epochs with mini-batch size of 8. Training took around 11 hours on RTX A4000 GPU. Figure 4(a) shows mAP at 0.5 IOU with respect to number of epochs. The model gets properly trained nearly after 500 epochs. Figure 4(b) shows mAP of 0.52 in the range of IOU varying from 0.5 to 0.95 indicating the accuracy of localizing and detecting targets at higher IOUs.

Table 1 Result with one stage detector: YOLOv7

Image Input Size	mAP@0.5	Recall@0.5	Precision@0.5	Prediction time
608 x 608	78%	0.7	0.85	8.2ms

Table 1 shows the performance parameters of YOLOv7 model on ship target dataset. 800X800 image is re-scaled to 608X608 by YOLOv7 model.

YOLOv7 vs YOLOv4: Both YOLOv7 & YOLOv4 models are trained on NVIDIA RTX A4000 GPU with HRSID and custom dataset of ship with 3000 training, 1647 validation and 300 testing images. The training time for YOLOv7 and YOLOv4 is around 16 & 11 hours respectively for 500 epochs. YOLOv7 takes less training time compared YOLOv4. mAP@0.5 and inference time of YOLOv4 are 57.9% and 11.6 ms respectively. mAP@0.5 and inference time of YOLOv7 are 78% and 8.2 ms respectively.

TESTING OF THE MODEL WITH TEST IMAGES

Figure 5 shows the detected targets on few images of test dataset. Ships detected are enclosed in bounding boxes along with confidence score. The size of input images given to the prediction model is 800 x 800 pixels. The model re-scales image to 608X608.

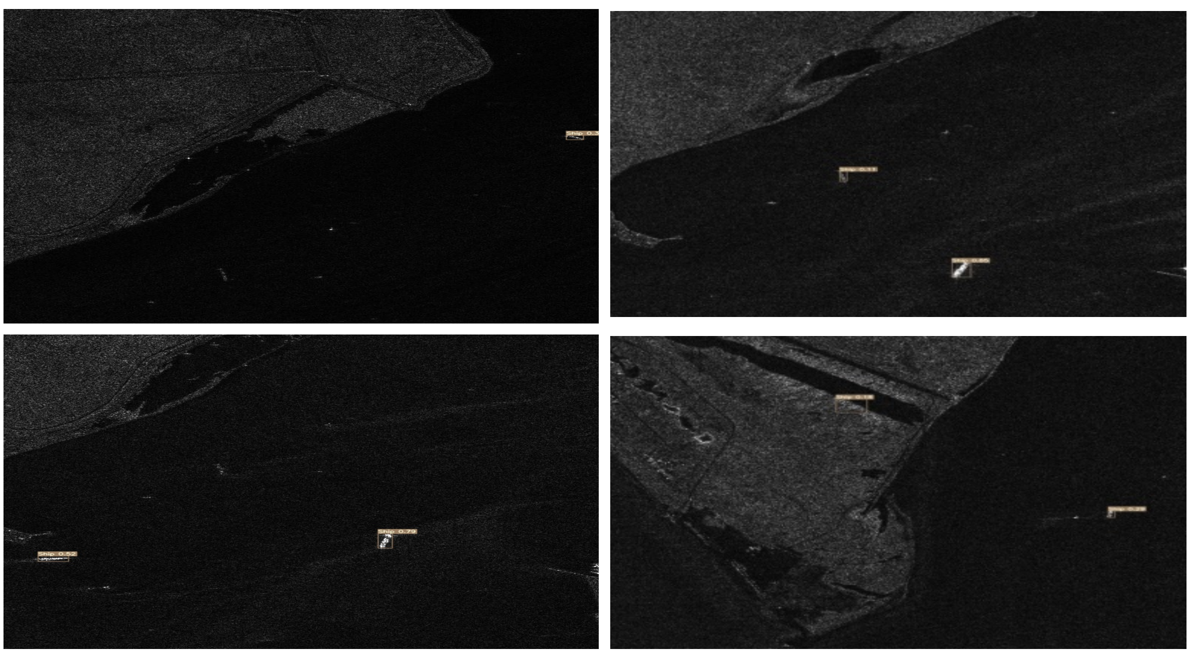

Figure 5: Detection of ships on test images (Detected ships localized with rectangular boxes)

6. CONCLUSION

In this paper, we have implemented a Real Time Target Detection on SAR images. SAR images cover large area for all day and for all weather conditions compared to optical images. YOLOV7 model is used for training and inference as it is superior over other YOLO versions in terms of accuracy and speed. Training is done with open source HRSID & custom dataset of ship as target. 78% of, mAP achieved with 0.5 IOU and inference time of model is 8.2 ms.

The model is trained on NIVIDA RTX A4000 GPU. Training time for 500 epochs is approximately 11 hours

There is a room for improvement in accuracy by varying CNN layers, prediction time may be minimized further on SAR by using better dataset and fine tuning of the model. YOLOv7 has good performance over YOLOv4 in terms of mAP and inference and training time. YOLOv7 can be trained on custom dataset without pre-trained weights. mAP further can be improved by training for more number of epochs and using better dataset. This model helps in detection of multiple targets in real time.

REFERENCES

1. El-Rabbany, Ahmed. Introduction to GPS: the global positioning system. Artech house, 2002.
2. Buck, G., and J. Gustincic. "Resolution limitations of a finite aperture." IEEE Transactions on Antennas and Propagation 15.3 (1967): 376-381.
3. Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv preprint arXiv:2207.02696 (2022).
4. Vo, Nguyen D., et al. "Empirical Study of the Performance of Object Detection Methods on Road Marking Dataset." 2022 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE, 2022.
5. Liu, Xiaoxu, and Wei Qi Yan. "Vehicle-Related Distance Estimation Using Customized YOLOv7." Image and Vision Computing: 37th International Conference, IVCNZ 2022, Auckland, New Zealand, November 24–25, 2022, Revised Selected Papers. Cham: Springer Nature Switzerland, 2023.
6. Gallo, Ignazio, et al. "Deep Object Detection of Crop Weeds: Performance of YOLOv7 on a Real Case Dataset from UAV Images." Remote Sensing 15.2 (2023): 539.
7. Jin, Lu, et al. "Real-time Fault Diagnosis of Photovoltaic Modules for Integrated Energy Systems Based on YOLOv7." (2023).
8. Liu, Wei, et al. "Ssd: Single shot multibox detector." Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016.
9. Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.
10. Thuan, Do. "Evolution of Yolo algorithm and Yolov5: The State-of-the-Art object detention algorithm." (2021).
11. Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "Scaled-yolov4: Scaling cross stage partial network." Proceedings of the IEEE/cvf conference on computer vision and pattern recognition. 2021.
12. Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014.
13. Hoiem, Derek, Santosh K. Divvala, and James H. Hays. "Pascal VOC 2008 challenge." World Literature Today 24 (2009).
14. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.