Size Invariant Ship Detection from SAR Images using YOLOv3 and Mask-RCNN

Unzela Inamdar[1], Pratik Kakade[2], Ameya Kale[3], Rutuja Jagtap[4], Prof. Nagesh Jadhav [5]

[1-4]Student, Dept. of Computer Science and Engineering, MIT School of Engineering, MIT ADT University, Pune, Maharashtra, India
[5]Professor, Dept. of Computer Science and Engineering, MIT School of Engineering, MIT ADT University, Pune, Maharashtra, India

---

Abstract - *Synthetic Aperture Radar (SAR) imagery has been used as a knowledge source for monitoring maritime activities, and its application for ship detection has been the main target of the many previous research studies. Here we have proposed an enhanced GPU based deep learning method for size invariant Ship Detection from SAR Images, i.e. to detect the ships from SAR Images, find its size (Small, Medium, or Large), and determine whether it is moving or not.*

The YOLO v3 deep learning framework is proposed to model the architecture and training model. This model is a real-time object detection system which performs a faster region-based convolutional network and single-shot multibox detector method.

There is minimal use of computational resources with increased accuracy in detection.

***Key Words*: Synthetic Aperture Radar (SAR) images, Ship detection, YOLO-v3, Deep Learning, Mask R-CNN.**

1. INTRODUCTION

Ship detection from remote sensing imagery has been a major application for maritime security. When talking about maritime security, we have to consider many things like traffic surveillance, protection against illegal fisheries, oil discharge control, and sea pollution monitoring.

Automated Identification System (AIS) is very effective at monitoring ships which are legally required to install a VHF transponder but fail to detect those which are not and those which disconnect their transponder.

So how do you detect these uncooperative ships?

SAR images are considered the most suitable sensors for object detection in space technology. It captures a wide surface of the environment, regardless of whether or time of day and flight altitude. Hence it has very high-resolution capabilities and gives high-quality images.

SAR has various applications in remote sensing and mapping of different surfaces of the Earth. It can be used in oceanology, glaciology, biomass, volcanoes, forestry, etc.

Before deep learning evolved traditional methods of target detection were divided into region selection eg. SIFT-scale invariant feature transform, HOG-histogram of an oriented gradient, and classifiers like SVM (Support Vector Machine) and Adaboost [11]. Unlike the sliding window and regional proposal based approach, YOLO sees the whole image during the training and testing period and thus encodes contextual information about classes and their appearance [11]. Further, YOLO makes less number of background errors than Fast R-CNN. But we have used Mask RCNN for instance segmentation.

Another most common approach is the CFAR-constant false rate use to detect targets with threshold with pixel's amplitude hence it is difficult to extract features [10].In addition, these methods are typically dependent on the statistical distribution of sea clutter [12,13,14], leading to poor robustness for new SAR imagery.

YOLO, predicts the bounding box and object class probability from the complete image in a single estimate.

We have to build a program to automatically identify whether a remotely sensed target is a ship or not. The program has to be extremely accurate because lives and billions of dollars in energy infrastructure are at stake. We will start by collecting huge data of satellite images of ships from various heights, (Roughly about 30 GB). To make the computations easy, we will mask some of the images, where there are ships present in the picture. While masking, we will make sure that all or almost all possibilities of the ships get covered. We will call this set the "**Ground Truth**". A suitable model will then be chosen in order to educate our program to identify a ship in the image in the dataset by comparing it with the ground truth. Once a ship is detected we will bound

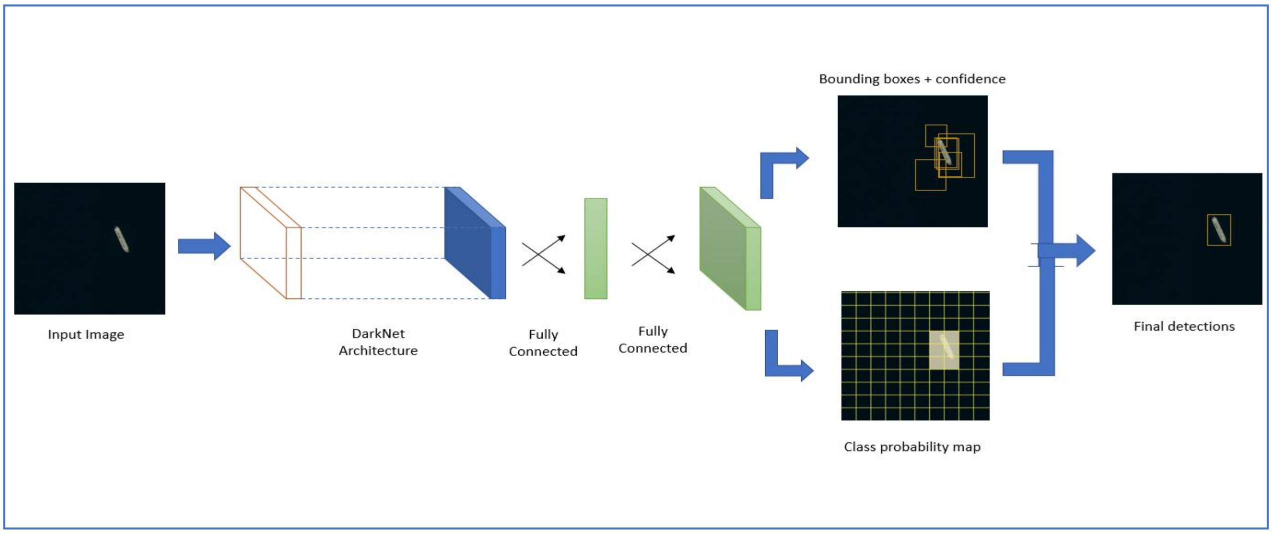


Fig – 1: YOLO block diagram on satellite image containing ships

It with the help of a bounding box. The bounding box will help us distinguish the ship's size as small, medium or large.

2. METHODOLOGY

2.1 Yolo v3:

Yolo v3 is the third iteration of the You Only Look Once (YOLO) algorithm which is typically used for object detection. Right from the beginning of its development there have been many changes in the architecture of the algorithm. This has resulted in faster detection speed and improved accuracy of detection.

Yolo uses the darknet architecture as a backbone for detections. Particularly Yolov3 uses the darknet -53 architecture for detections. As the name suggests it consists of 53 convolutional layers which helps to improve its accuracy over the previous version which used the darknet-19 architecture.

Fig.1 shows the process of detection of ships in satellite imagery. Multiple bounding boxes are formed on image with different confidence. Most significant region dependent on the confidence of these boxes is calculated. This provides us with most significant regions where the object maybe located. This can be imagined as different regions representing different classes of objects.

2.2 Mask RCNN:

The Mask RCNN framework was created by Facebook's AI Research team or FAIR in 2017. This relatively new Framework is an extension of Faster RCNN. So, just like Fast RCNN and Faster RCNN, Mask RCNN is also a deep neural network. Mask RCNN solves the problems of instance segmentation in machine learning and computer vision. [15]

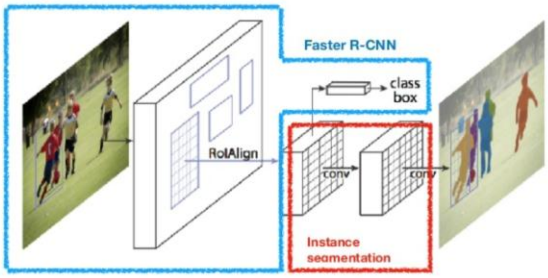
The process of Mask RCNN can be broken down into two steps:

1. Generation of potential regions of interest using RPN and RoI Align and also using the Ground Truths.
2. Prediction of the class of the object, the bounding box and the mask in the pixel level. [15]

Working of Fast RCNN in short:

1. Feature Extraction: Faster RCNN makes use of CovNet for generating feature maps of the image.
2. Propose Potential Regions: The feature maps generated from CovNet are further passed to the RPN (Region Proposal Network) where the bounding boxes over the regions are returned.
3. Making it uniform: These regions and bounding boxes are then passed to RoI (Region of Interest) pooling and RoI Align where they are brought down to the same size. This helps further makes computations easier and faster.
4. FC Layers: Fully Connected layers or FC layers are the final steps in Faster RCNN, where the proposals are passed, and classification takes place. Outputs are bounding boxes over the objects. [16]

Now, let's understand the working of Mask RCNN

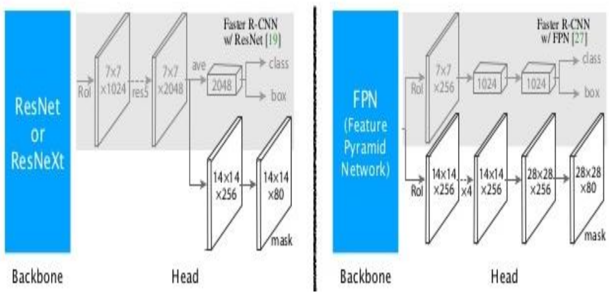


[19]

Fig – 2: Mask RCNN

2.3 Backbone Model:

The backbone of Mask RCNN is the ResNet 101 architecture. ResNet 101 to Mask RCNN is the same as what CovNet is to Faster RCNN. ResNet 101 extracts features from images and generates feature maps. A FPN or Feature Pyramid Network is formed with the help of these feature maps. These feature maps and the FPN are then passed over to the next layer. [16]



[19]

Fig – 3: Backbone Model

2.3 Region Proposal Network (RPN):

The Region Proposal Network or RPN tries to predict the objects present in the image using the feature maps and FPN received from the previous layers. Once a potential object is found, it then draws a bounding box around it. [16]

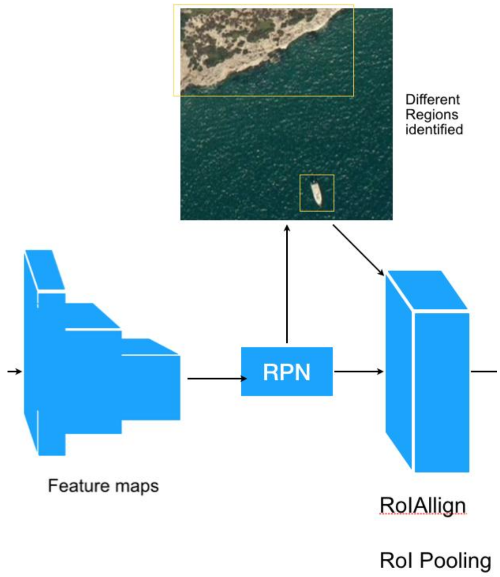


Fig – 4: Region Proposal Network (RPN)

2.4 Region of Interest (RoI):

The identified objects or regions are then passed over to the Region of Interest layer. This layer takes care of different shapes of the regions by applying a pooling layer and converting them into the same shapes. This further helps computation. [16]

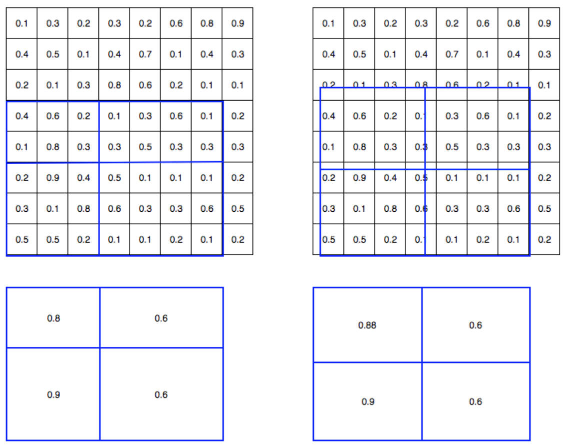


Fig – 5: Region of Interest (RoI)

In the above figure, you can see the difference between RoI and RoI Align. In the left column, the target features are forced to realign with the boundaries of the feature maps, hence producing different cell sizes. Whereas RoI Align allows every target cell to be of the same size. Mask RCNN makes use of RoI Align. [18]

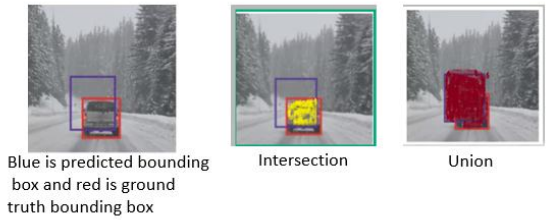
2.5 The segmentation mask:

After computing the RoI, we have to compute the IoU over all of the predicted regions. IoU stands for Intersection over Union and is calculated with the help of ground truths.

IoU = Area of the intersection / Area of the union [16]

When the value of IoU equals 1, it implies that the predicted boxes overlap perfectly with the ground truth boxes.

To make things a little less rigid, Non-Max Suppression is applied, where we basically consider all the predicted boxes with IoU > 0.5. The rest of the boxes are removed. In case of same objects, it will choose the box with the highest value of IoU and will discard the rest.



[17]

Fig – 6: The segmentation mask

This completes the process of Mask RCNN, where we get the masks for the objects in the image.

Mask RCNN takes a lot of time to train. On average it takes around a couple of days to be completely trained. Therefore, we took help from the pretrained weights of the COCO dataset trained on the Mask RCNN model.

3. SYSTEM DIAGRAM

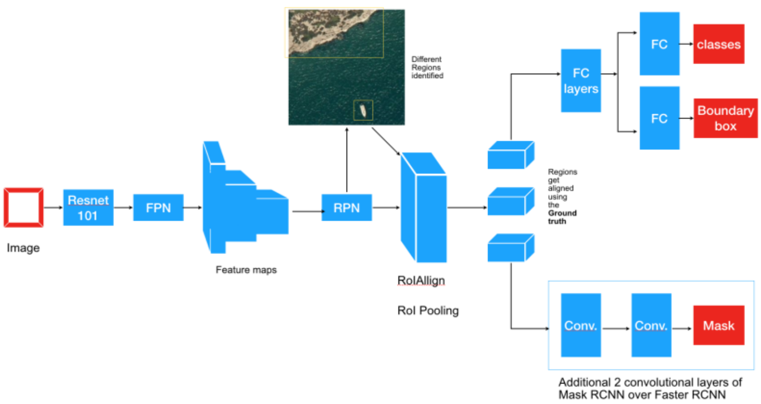
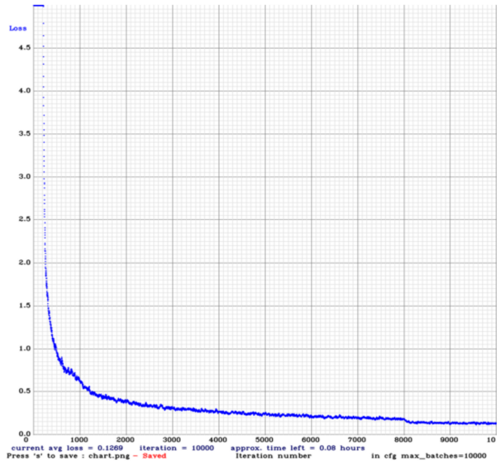


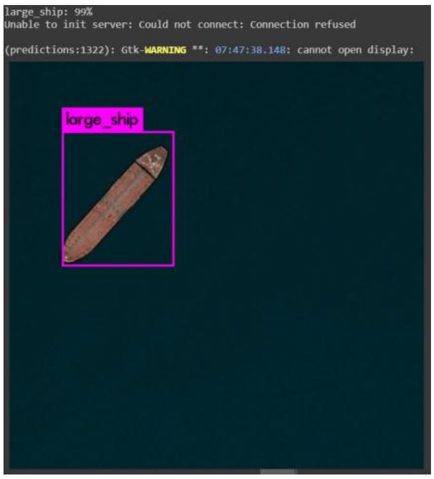
Fig – 7: Mask R-CNN

4. EXPERIMENTAL RESULTS

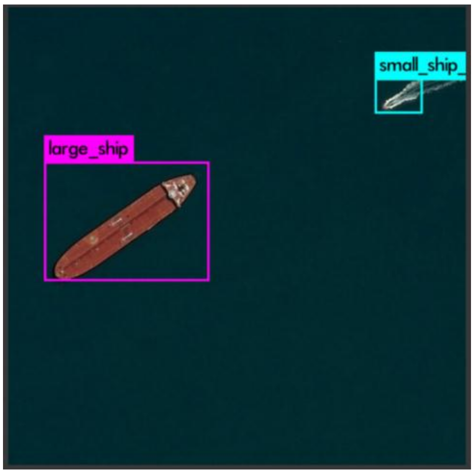
For Yolo v3 :



Loss Graph Plot :
x-axis = No. of iterations
y-axis = Loss

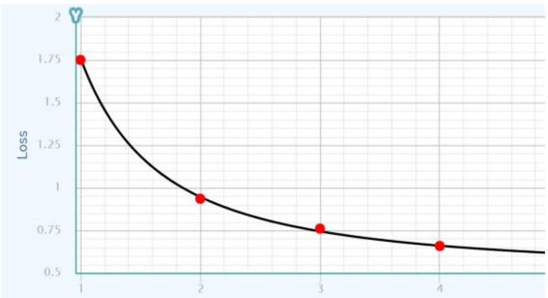


Confidence: 99%
Total Execution Time: 40.67ms
Computation Resources used: 65.33 BFLOPS

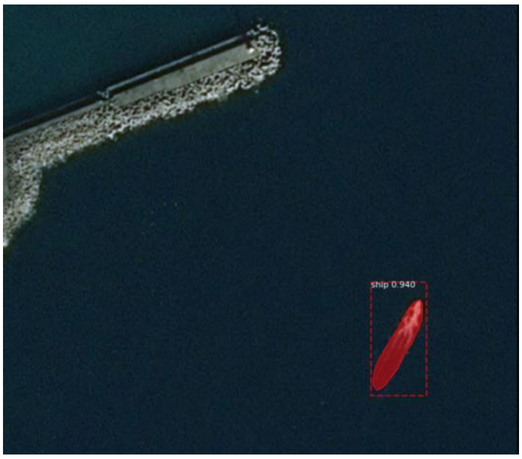


Confidence:
Large Ship = 99%
Small Ship Moving = 35%
Total Execution Time: 40.50ms
Computation Resources used: 65.33 BFLOPS

For Mask RCNN:



Loss graph for 5 epochs



Accuracy: 94%

Multiple ships are also detected

5. CONCLUSIONS

Our proposed system works efficiently not only on a single ship image but also on images that have multiple ships.

Hence, we were successfully able to detect the ships from SAR Images, find its size (Small, Medium, or Large), and determine whether it is moving or not.

FUTURE ENHANCEMENT

Its scope can further be extended and used for applications like:

- Oil spill detection.
- Enemy ship detection.
- Sea pollution monitoring.
- Illegal fishery detection.

REFERENCES

[1] Liu, Y., Ma, H., Yang, Y., & Liu, Y. (2011). Automatic ship detection from SAR images. Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services

[2] Arivazhagan, S., Lilly Jebarani, W. S., Newlin Shebiah, R., Ligi, S. V., Hareesh Kumar, P. V., & Anilkumar, K. (2019). Significance based Ship Detection from SAR Imagery. 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT).

[3] Iervolino, P., Guida, R., Lumsdon, P., Janoth, J., Clift, M., Minchella, A., & Bianco, P. (2017). Ship detection in SAR imagery: A comparison study. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

[4] Zhai, L., Li, Y., & Su, Y. (2016). Segmentation-based ship detection in harbor for SAR images. (2016). CIE International Conference on Radar (RADAR).

[5] Ao, W., & Xu, F. (2018). Robust Ship Detection in SAR Images from Complex Background. 2018 IEEE International Conference on Computational Electromagnetics (ICCEM).

[6] Li, L., & Wang, J. (2012). SAR image ship detection based on Ant Colony Optimization. 2012 5th International Congress on Image and Signal Processing.

[7] Nan, J., Wang, C., Zhang, B., Wu, F., Zhang, H., & Tang, Y. (2013). Ship wake CFAR detection algorithm in SAR images based on length normalized scan. 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS.

[8] Hou, B., Yang, W., Wang, S., & Hou, X. (2013). SAR image ship detection based on visual attention model. 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS.

[9] Liu, L., Chen, G., Pan, Z., Lei, B., & An, Q. (2018). Inshore Ship Detection in Sar Images Based on Deep Neural Networks. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.

[10] M. Kang, X. Leng, Z. Lin and K. Ji. (2017). "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai.

[11] Chang, Y.-L.; Anagaw, A.; Chang, L.; Wang, Y.C.; Hsiao, C.-Y.; Lee, W.-H. (2019). Ship Detection Based on YOLOv2 for SAR Imagery.

[12] El-Darymli, K.; McGuire, P.; Power, D.; Moloney, C.R. (2013). Target detection in synthetic aperture radar imagery: A state-of-the-art survey.

[13] El-Darymli, K.; Gill, E.W.; Mcguire, P.; Power, D.; Moloney, C. (2016). Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review.

[14]Huang, X.; Yang, W.; Zhang, H.; Xia, G.-S. (2015). Automatic Ship Detection in SAR Images Using Multi-Scale Heterogeneities and an A Contrario Decision. Remote Sens.

[15]Xiang Zhang. Simple Understanding of Mask RCNN. (2018). https://medium.com/@alittlepain833/simple-understanding-of-mask-rcnn-134b5b330e95 .

[16]Pulkit Sharma. Computer Vision Tutorial: Implementing Mask R-CNN for Image Segmentation (with Python Code). (2019). https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/.

[17] Renu Khandelwal. Computer Vision: Instance Segmentation with Mask R-CNN. (2019). https://towardsdatascience.com/computer-vision-instance-segmentation-with-mask-r-cnn-7983502fcad1.

[18] Jonathan Hui. Image Segmentation with Mask RCNN. (2018). https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272.

[19]Chanuk Lim. Mask RCNN. https://www.slideshare.net/windmdk/mask-rcnn.