Real Time Detection And Segmentation Of Ships In Satellite Images

Ramani S, Prabakaran N, Kannadasan R, Rajkumar S*

Abstract:— The goal of this work is to build an algorithm to automatically detect and segment ships in satellite images. A number of factors can make ship detection and segmentation a challenging task. Some of these factors include flaws in image quality like uneven brightness, obstruction, and also the fact that there are many images which have similar shape, color and texture. Another problem is that objects like islands, ports, whales etc look quite similar to ships. The algorithm had to be extremely accurate because lives and billions of dollars in energy infrastructure is at stake. In this paper we have used a custom MASK R CNN with backbone as ResNet 50 trained using MS COCO weights. Using mask shape of size 14*14, the algorithm is able to detect and segment ships with a Mean Average Precision of 0.605. The model takes approximately 30 seconds to segment 30 images. Hence this model can be used for real time maritime applications.

Index Terms:— Mask R CNN, Instance segmentation, MS Coco, GPU, Ship detection and segmentation

1 INTRODUCTION

Hundreds of ships are constantly orbiting the earth. These ships are used for Global Positioning System (GPS) applications, ship detection, communications and forecasting weather among other applications. Ship detection from satellite images is an important application for maritime applications like ship traffic surveillance, protection against illegal fisheries, oil discharge control and sea pollution monitoring. This is currently manually being done through the use of an Automated Identification System (AIS), which uses Very High Frequency (VHF) radio frequencies. These frequencies wirelessly broadcast the information for the location of ships, destination and give identity to nearby receiver devices on other ships and land-based systems. AIS are very effective at monitoring ships which are legally required to install a VHF transponder, but fail to detect those which are not. Also in some cases people disconnect their transponder which is a quite common case. Satellite imagery can be used in these cases as it removes the human element out of the loop. Synthetic Aperture Radar (SAR) imagery uses radio waves to image the Earth's surface. Unlike optical imagery, the wavelengths which the instruments use are not affected by the time of day or meteorological conditions, enabling imagery to be obtained day or night, with cloudy, or clear skies. Hundreds of satellites orbiting the earth are collecting these images constantly which could be used to make algorithms for ship detection and segmentation. This can be of great help to the maritime security and offshore operations in the surveillance, energy and military sector. There are four major types of problems in computer vision which are summarized below. Image Classification: Classify the object present in the image. Object Detection: Identify and draw bounding boxes for every object within an image. Semantic Segmentation: Categorize objects in each pixel for every object within an image. Instance Segmentation: Categorize object instance in each pixel for every object within an image.

- *Ramani S is working as Assistant Professor(Senior) in Vellore Institute of Technology, India, E-mail: ramani.s@vit.ac.in*
- *Prabakaran N is working as Assistant Professor(Senior) in Vellore Institute of Technology, India, E-mail: prabakaran.n@vit.ac.in*
- *Kannadasan R is working as Assistant Professor(Senior) in Vellore Institute of Technology, India, E-mail: kannadasan.r@vit.ac.in*
- *Rajkumar S is working as Associate Professor in Vellore Institute of Technology, India, E-mail:rajkumars@vit.ac.in*

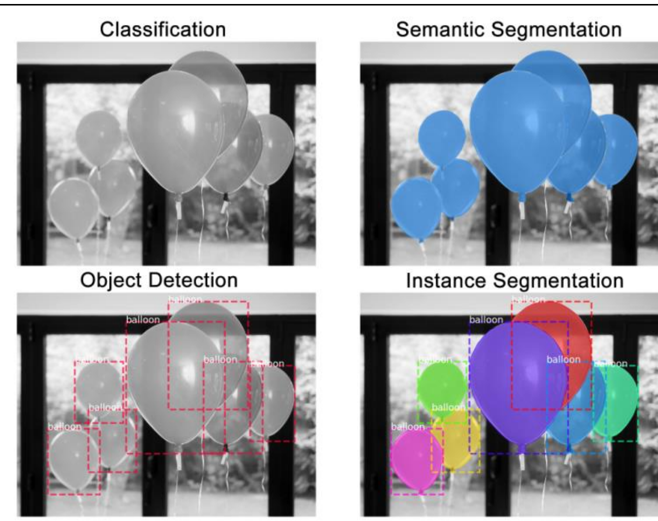

***Fig. 1**. Classification, Object detection, Semantic segmentation, Instance Segmentation*

2.1 Mask R-CNN Architecture

In this section, a brief overview of MASK R-CNN architecture and how it has been proven to be the state of art in instance segmentation problems is presented.

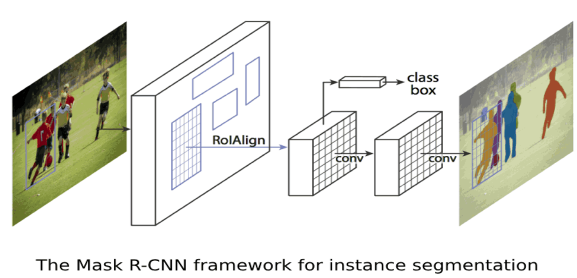

***Fig. 2**. Mask R-CNN framework for Segmentation*

The earlier state of art R-CNN, Fast R-CNN and Faster R-CNN had two outputs for each candidate object, a class label and a bounding-box offset. MASK R-CNN extends the approach by adding a third branch that outputs the object mask. The additional mask output is distinct from the class and box outputs, requiring extraction of much finer spatial layout of an object. The details from the original paper can be summarized in the following four steps:

1. Backbone model: A standard convolutional neural network that serves as a feature extractor.
2. Region Proposal Network (RPN): The RPN scans each region and predicts whether an object is present or not using the regions defined by the anchor boxes.
3. Region of Interest Classification: The algorithm takes the regions of interest proposed by the RPN as inputs and outputs a classification for objects and a bounding box using softmax layer.
4. Segmentation Masks: The positive ROI regions are taken as inputs and pixel masks with float values are generated as outputs for the objects detected.

2 EXISTING SYSTEM

A lot of existing work related to our research has already taken place. In this section, we briefly discuss some of the existing work: Machine learning and deep learning have attracted increasing attention and achieved great success in different areas including Computer Vision (Krizhevsky et al). Many significant breakthroughs have been made in the area of image classification, such as AlexNet (Krizhevsky et al). (Ying et al) has compared traditional approach using algorithms like linear discriminant analysis, K nearest neighbors, naive Bayes and random forest vs. CNN based for training the model. (Liu et al) has done a comparative study using variants of feature pyramid network using rotating bounding box, non-maximum suppression and Region of interest (ROI) align method. (Kang et al) proposed a ship detection method using pre trained VGG16 as FASTER R-CNN framework. (Zhu et al) in his paper proposed a hierarchical complete and operational approach based on shape and texture features, which is considered a sequential coarse-to-fine elimination process of false alarms. (Yu et al) proposed a high resolution optical remote sensing dataset for ship recognition. They analyzed several unique features for the dataset including ship hierarchy, bounding information from the labeled dataset, unbiasedness and rich helping tools. (Kaiming et al) proposed a conceptually simple, flexible, and general framework for object instance segmentation. The model named MASK-RCNN both efficiently detects objects in the image and simultaneously generating a high-quality segmentation mask for each instance of object.

2.1 Data Set

In this work, a public dataset from Kaggle on the Airbus Ship Detection Challenge is obtained. The dataset contains more than 100 thousand 768 × 768 images taken from satellite. The total size of the dataset is more than 30 Gb. Along with the images in the dataset, is a CSV file that lists all the images ids and their corresponding pixels coordinates. These coordinates represent segmentation bounding boxes of ships. Not having pixel coordinates for an image means that particular image doesn't have any ships.

Fig. 3. *Sample Image*

3 PROPOSED SYSTEM

The original parameter values from MASK R-CNN paper are summarized in the below table. We took these parameter values as reference for our work.

TABLE 1

ORIGINAL PARAMETERS OF MASK R-CNN

Parameters	*Values*
train validation ratio	*0.2*
image width	*768*
image height	*768*
gpu count	*1*
images per gpu	*2*
number of classes	*2*
steps per epoch	*100*
validation steps	*100*
minimum detection confidence	*0.95*
detection nms threshold	*0.05*

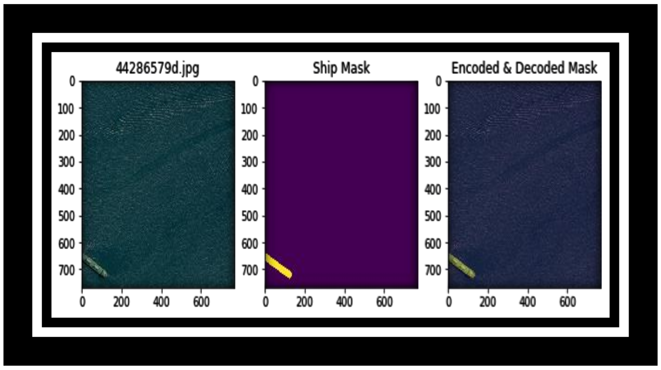

Fig. 4*. sample image, ship mask, encoded and decoded mask*

The dataset is split into training and validation set with 80 % and 20% images respectively. Function is created for encoding and decoding mask on top of every image then Mask R-CNN for segmenting the ships in image with a confidence score between 0 and 1 is set. The time required to mask out ships in the complete dataset was 9.82 second. Pre trained MS COCO weights are used for training the model. Since these weights have already been trained on a large variety of objects hence they provide a good place to start learning from.

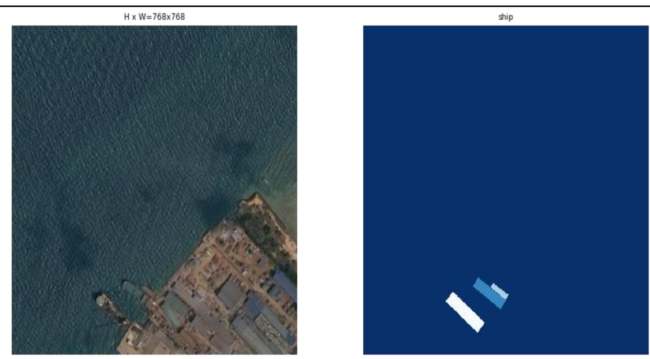

Fig. 5*. Sample Image, Masked Image*

Various hyper-parameters like learning momentum, learning rate, weight decay, backbone, pool size, mask pool size and mask shape are changed and their effect on time to inference is studied and mean average precision which are summarized next.

4 Experimental Results

It is found that using mask shape of size 14*14 and Resnet50 backbone, the algorithm is able to detect and segment ships with a mean average precision of 0.605. It took the model approximately 30 seconds to segment 30 images. Hence this model can be used for real time maritime applications.

TABLE 2

COMPARING TIME FOR INFERENCE AND MEAN AVERAGE PRECISION WITH CHANGING PARAMETERS

Parameters	*Value*	*Time for inference(min)*	*Mean Average Precision*
learning momentum	*0.99*	*0.33*	*0.392*
learning rate	*0.0001*	*0.32*	*0.472*
weight decay	*0.00001*	*0.32*	*0.471*
backbone	*resnet50*	*0.33*	*0.572*
pool size	*14*	*0.33*	*0.415*
mask pool size	*28*	*0.33*	*0.475*
mask shape	*14,14*	*0.33*	*0.605*

The comparison of original parameters with the modified parameters is done. The changes made are summarized in the below table.

Parameters	*Original*	*Proposed Parameters*
Backbone	*ResNet101*	*Resnet50*
Learning momentum	*0.9*	*0.9*
Learning rate	*0.001*	*0.001*
Weight decay	*0.0001*	*0.0001*
Pool size	*7*	*7*
Mask size	*14*	*14*
Mask shape	*28*28*	*14*14*

The Mean Average Precision (MAP) value using original MASK R-CNN parameters vs our modified/ proposed parameters is compared. Using original parameters MAP value was found to be 0.472 while using our custom parameters the value obtained was 0.605.

	Original	*Proposed*
Mean average precision	*0.472*	*0.605*

There are two loss terms associated while training the model - RPN loss which is the loss associated with detecting boxes that will be examined by a classifier and regressor to check the occurrence of objects and MASK R-CNN class loss which is the loss associated with detecting and segmenting individual classes present in the image. The RPN loss and MASK R-CNN class loss are plotted below as a function of epochs. We trained the model for 5 epochs as our focus is on using this model on real time applications.

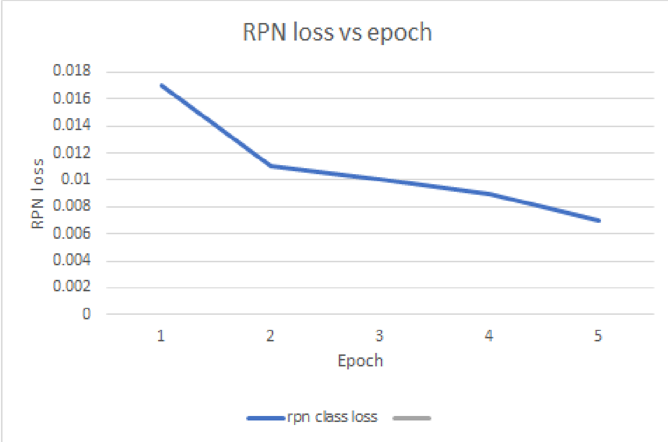

***Fig. 6**. RPN loss vs Epochs*

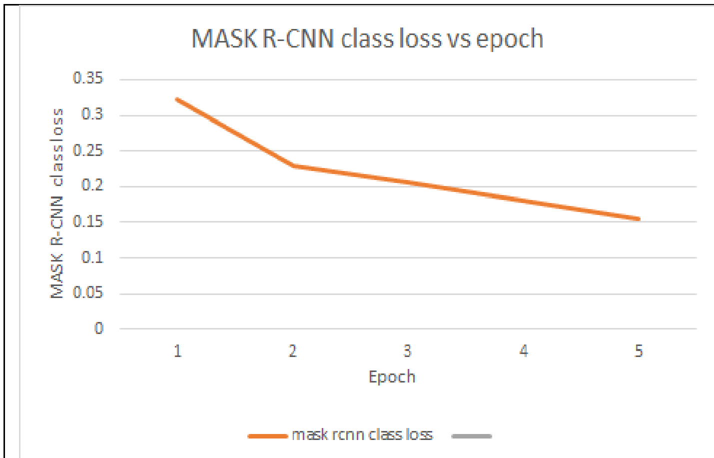

***Fig. 7**. MASK R-CNN vs Epochs*

The final results of detecting and segmenting ships in satellite images with confidence scores is presented.

***Fig. 8**. Detection and Segmentation with confidence score*

5 Conclusion

In this paper we have combined both ship detection and segmentation in a single problem and solved it using MASK R-CNN architecture. We have used MS-COCO weights for training the model. We changed the backbone from ResNet101 to ResNet50 and mask shape from 28*28 to 14*14 compared to the original paper. The model was able to to segment ships in satellite images with a MAP value of 0.605. It took 30 seconds for the model to segment 30 ships. Hence this method can be used in real time maritime applications for logistics and transportation team in northern countries like Sweden, Norway and Canada. It could bring a whole new dimension of transport for container ships and vessels by tracking ships from satellite images in real time

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[2] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[3] Kaggle. Dataset For Airbus Ship Detection Challenge. https://www.kaggle.com/c/airbus-ship-detection/ data, 2018.

[4] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and FeiFei, L. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.

[5] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12, pp. 1097–1105, USA, 2012. Curran Associates Inc. URL http://dl.acm.org/citation.cfm?id=2999134.2999257.

[6] Ying Chen, Junwen Zheng, Zhengqing Zhou. Airbus Ship Detection - Traditional v.s. Convolutional Neural Network Approach

[7] Cai, Z.; Fan, Q.; Feris, R.S.; Vasconcelos, N. A unified multi-scale deep convolutional neural network for fast object detection. In Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 354–370.

[8] He, K., Gkioxari, G., Dollár, P., & Girshick, R.B. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2980-2988.

[9] Liu, Y.; Zhang, M.H.; Xu, P.; Guo, Z. W. SAR ship detection using sea-land segmentation-based convolutional neural network. International Workshop on Remote Sensing with Intelligent Processing IEEE. Shanghai, China, 18-21, May 2017; pp. 1-4.

[10] Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. International Workshop on Remote Sensing with Intelligent Processing IEEE. China, 18-21, May 2017; pp. 1-4.

[11] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1137–1149.

[12] Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.

[13] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[14] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[15] Zhu, C., Zhou, H., Wang, R., and Guo, J. (2010). A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. IEEE Trans. Geosci. Remote Sens., 48(9):3446–3456.
[16] Yu, Y., Guan, H., and Ji, Z. (2015). Rotation-invariant object detection in high-resolution satellite imagery using superpixel-based deep hough forests. IEEE Geosci. Remote Sens. Lett., 12(11):2183–2187
[17] Yang, G., Li, B., Ji, S., Gao, F., and Xu, Q. (2014). Ship detection from optical satellite images based on sea surface analysis. IEEE Geosci. Remote Sens. Lett., 11(3):641–645.
[18] Liu, G., Zhang, Y., Zheng, X., Sun, X., Fu, K., and Wang, H. (2014). A new method on inshore ship detection in high-resolution satellite images using shape and context information. IEEE Geosci. Remote Sens. Lett., 11(3):617–621.
[19] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pages 580–587.