International Journal of Advanced Development in Science and Technology
IJADST Volume : 3 Issue : 03 April 2021 www.ijadst.com e-ISSN: 2582-1059

SHIP DETECTION FROM HIGH-RESOLUTION REMOTE SENSING IMAGE USING MASK R-CNN

[1] Suganya T, [2] A. Josephine Anitha,
[1]Research Scholar, [2] HOD,
[1, 2] Ethiraj College for Women, Chennai
[1] Thangasugan94@gmail.com, [2] ajan_jose@yahoo.com.

ABSTRACT:

High resolution satellite image processing is one of the most growing fields in research today. There is so much to explore and so many ways to do it that it seems full of endless opportunities and possibilities. There are several features which can be extracted like buildings, roads etc. from land satellite images and ships, boats etc. from satellite images of sea and ocean. In this paper we will be concentrating on detecting ships automatically from the images obtained by various satellites. This is one of the major challenging tasks due to various disturbances and noises in these kinds of images. Ships can be found in different sizes as well as shapes which make it more difficult to find a pattern or some regularity in these images.

It is comparatively easier in homogeneous environment consisting of just ships of different types in water. But when it comes to heterogeneous environment consisting of other elements like coasts, harbor, vessel, rocks, islands etc. the challenge increases tenfold. There are various statistical and image processing approaches which can do this manually then again this won't be that efficient, changing the parameters again and again with different images and all can be a time consuming and tedious process. That's why we choose one of the modern approaches which provide us with the opportunity of being automatic or at least semi-automatic, deep learning.

Deep learning together with computer vision opens doors to the possibilities which we couldn't even have thought of. In this paper we will be exploring those possibilities with the help of a certain Convolutional neural network known as Mask R_CNN and implementing it using transfer learning. This algorithm gives very high accuracy in classification of satellite images without doing any manual extraction and works with complex heterogeneous backgrounds too.

Keywords: Synthetic Aperture Radar (SAR) images, Ship detection, Deep Learning, CNN, and 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. 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 it with the help of a bounding box.

International Journal of Advanced Development in Science and Technology
IJADST Volume : 3 Issue : 03 April 2021 www.ijadst.com e-ISSN: 2582-1059

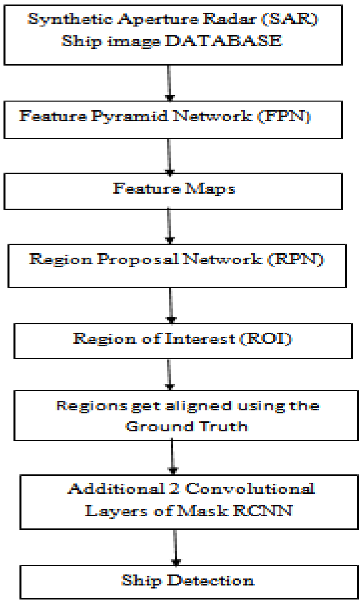

Fig 1 steps for ship detection

The dataset consists of image chips extracted from Planet satellite imagery collected over the San Francisco Bay and San Pedro Bay areas of California. It includes 4000 80x80 RGB images labeled with either a "ship" or "no-ship" classification. Image chips were derived from Planet Scope full-frame visual scene products, which are orthorectified to a 3 meter pixel size.

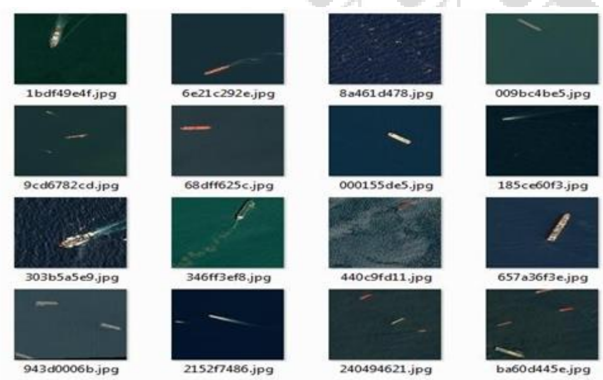

Fig 2 Synthetic Aperture Radar (SAR) Ship image data set

2. PROPOSED SYSTEM

This end-to-end system contains four sub-networks with different functions. The feature map of the input image is obtained by the Feature Pyramid Network (FPN) first, and then the scene mask of target and non-target area is extracted by the scene mask extraction network (SMEN). With the feature combination between the output of FPN and the estimated scene mask, the false alarm targets existing in non-target area are eliminated entirely. Then Region Proposal Network (RPN) uses the combined feature map to generate the proposed bounding boxes. 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. This completes the process of Mask RCNN, where we get the masks for the objects in the image. Therefore, we took help from the pretrained weights of the COCO dataset trained on the Mask RCNN model.

2.1. ADVANTAGES OF PROPOSED SYSTEM

- The false alarms caused by the onshore ship-like objects may decrease the accuracy and feasibility of these DCNN-based detection frameworks.
- Mask R-CNN, is proposed to reduce the onshore false alarms.
- This proposed system effectively able to extract all the features of an image.

3. MASK RCNN

The Mask RCNN framework was created by Face book'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

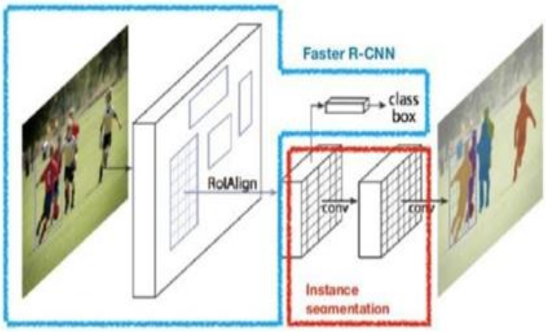

Fig 3: Mask RCNN

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

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

3.1. WORKING OF FAST RCNN IN SHORT

- **Feature Extraction:** Faster RCNN makes use of CovNet for generating feature maps of the image.
- **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.
- **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.
- **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.

4. 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

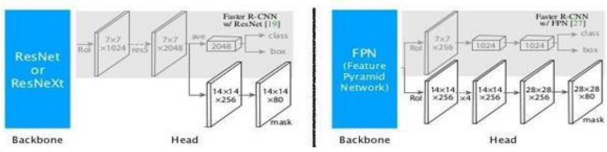

Fig 4: Backbone Model

4.1. 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

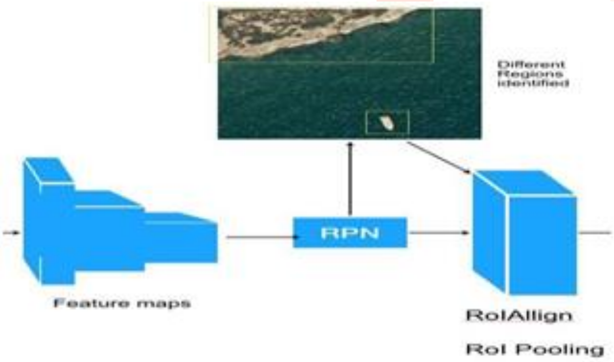

Fig 5: Region Proposal Network (RPN)

4.2. 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.

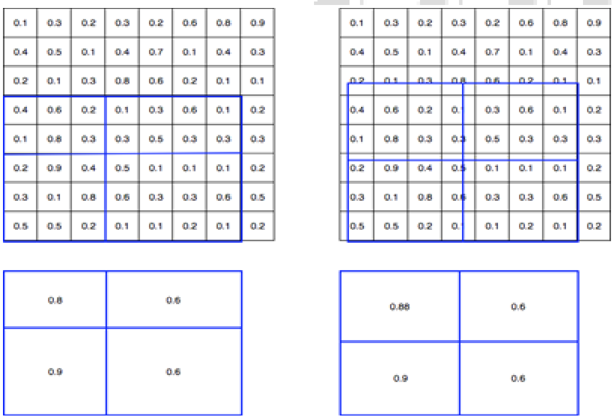

Fig 6: Region of Interest (RoI)

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.

4.3. 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.

$$IoU = Area\ of\ the\ intersection\ / \ Area\ of\ the\ union$$

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.

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

International Journal of Advanced Development in Science and Technology
IJADST Volume : 3 Issue : 03 April 2021 www.ijadst.com e-ISSN: 2582-1059

COCO dataset trained on the Mask RCNN model.

5. CONCLUSION

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. In the near shore ship detection mission, the existing DCNN-based detection methods pay little attention to the suppression of onshore false alarms.

In practical application, some onshore ship like objects, such as dock, roof, and road, is interpreted as targets of interest with high probability, even if the trained network has high detection accuracy. In fact, it is more suitable to detect targets after excluding the non-target area by using the scene information. In this proposal, an effective DCNN-based ship detection method, named as Scene Mask R-CNN, is proposed to reduce the onshore false alarms.

6. REFERENCES

[1]. G.E.Hinton, S.Osindero, and Y.-W.Teh, ‘‘A fast learning algorithm for deep belief nets,’’ Neural Comput., vol.18, no.7, pp.1527–1554, Jul.2006.

[2]. S.Ren, K.He, R.Girshick, and J. Sun, ‘‘Faster R-CNN: Towards real time object detection with region proposal networks, ’’IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[3]. W.Liu, D.Anguelov, D.Erhan, C.Szegedy, and S. Reed, ‘‘SSD: Single shot multibox detector,’’ in Proc. Eur. Conf. Comput. Vis. (ECCV), in Lecture Notes in Computer Science, vol. 9905. Amsterdam, the Netherlands: Springer, Oct. 2016, pp. 21–37.

[4]. Y.-L. Chang, A.Anagaw, L.Chang, Y.C.Wang, C.-Y.Hsiao, and W.H.Lee, ‘‘Ship detection based on YOLO v2 for SARimagery, ’’Remote Sens., vol. 11, no. 7, p. 789, Apr. 2019.

[5]. Z. Lin, K.Ji, X.Leng, and G.Kuang, ‘‘Squeeze and excitation rank faster R-CNN for ship detection in SAR images,’’ IEEE Geosci. Remote Sens. Lett., vol. 16, no. 5, pp. 751–755, May 2019.

[6]. C.Dong, J.Liu, and F.Xu, ‘‘Ship detection in optical remote sensing images based on saliency and a rotation-invariant descriptor,’’ Remote Sens., vol. 10, no. 3, p. 400, Mar. 2018.

[7]. T. Zhang and X. Zhang, ‘‘High-speed ship detection in SAR images based on a grid convolutional neural network,’’ Remote Sens., vol. 11, no. 10, p. 1206, May 2019.

[8]. X. Yang, J. Yang, J. Yan, Y. Zhang, T. Zhang, Z.Guo, S. Xian, and K. Fu, ‘‘SCRDet: Towards more robust detection for small, cluttered and rotated objects,’’ Aug. 2019, arXiv:1811.07126. [Online]. Available: https://arxiv.org/abs/1811.07126v4

[9]. L.Qu, S. Wang, N. Yang, L. Chen, L. Liu, X. Zhang, F. Gao, and J. Dong, ‘‘Improving object detection accuracy with region and regression based deep CNNs,’’ in Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC), Dec. 2017, pp. 318–323.

[10]. Q.Li,L.Mou, Q.Liu, Y.Wang, and X.X.Zhu, ‘‘HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery,’’ IEEE Trans. Geosci. Remote Sens., vol. 56, no. 12, pp. 7147–7161, Dec. 2018.

[11]. Z.Zhang, W.Guo, S.Zhu, and W. Yu, ‘‘toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks, ’’IEEE Geosci. Remote Sens. Lett., vol. 15, no. 11, pp. 1745–1749, Nov. 2018.

[12]. L.Gao,Y.He,X.Sun,X.Jia,andB.Zhang,‘‘ Incorporatingnegativesample training for ship detection based on deep learning,’’ Sensors, vol. 19, no. 3, p. 684, Feb. 2019.

[13]. X. Li, S. Wang, B. Jiang, and X. Chan, ‘‘Inshore ship detection in remote sensing images based on deep features,’’ in Proc. IEEE Int. Conf. Signal Process. Commun. Comput. (ICSPCC), Oct. 2017, pp. 1–5.

[14]. F. Wu, Z. Zhou, B. Wang, and J. Ma, ‘‘Inshore ship detection based on convolutional neural network in optical satellite images,’’ IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 11, pp. 4005–4015, Nov. 2018.

[15]. G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, ‘‘Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images,’’IEEEJ.Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 5, pp. 1633– 1644, May 2018.

[16]. V. Badrinarayanan, A. Kendall, and R. Cipolla, ‘‘SegNet: A deep convolutional encoder- decoder architecture for image segmentation,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[17]. O. Ronneberger, P. Fischer, and T. Brox, ‘‘U-net: Convolutional networks for biomedical image segmentation,’’ in Medical Image Computing and

Computer-Assisted Intervention (Lecture Notes in Computer Science), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234– 241.

[18]. R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, ‘‘Deep UNet: A deep fully convolutional network for pixel-level sea-land segmentation,’’ IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 11, pp. 3954–3962, Nov. 2018.

[19]. S.Sun,Z.Lu, W.Liu, W.Hu, and R.Li, ‘‘SH ipnetfor semantic segmentation on VHR maritime imagery, ’’inProc.Int.Geosci. RemoteSens. Symp. (IGARSS), Jul. 2018, pp. 6911– 6914.

[20]. H. Lin, Z. Shi, and Z. Zou, ‘‘Maritime semantic labeling of optical remote sensing images with multi-scale fully convolutional network,’’ Remote Sens., vol. 9, no. 5, p. 480, May 2017.

[21]. S. Nie, Z. Jiang, H. Zhang, B. Cai, and Y. Yao, ‘‘Inshore ship detection based on mask R- CNN,’’ in Proc. Int. Geosci. Remote Sens. Symp., Jul. 2018, pp. 693–696.

[22]. E. Shelhamer, J. Long, and T. Darrell, ‘‘fully convolutional networks for semantic segmentation,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, Apr. 2017.

[23]. J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, ‘‘Selective search for object recognition,’’ Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, Apr. 2013.

[24]. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770–778.

[25]. K. Simonyan and A. Zisserman, ‘‘Very deep convolutional networks for large-scale image recognition,’’ Sep. 2014, arXiv:1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556

[26]. M. D. Zeiler, G. W. Taylor, and R. Fergus, ‘‘Adaptive deconvolutional networks for midand high level feature learning, ’’inProc. IEEE Int.Conf. Comput. Vis., Nov. 2011, pp. 2018–2025.

[27]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, ‘‘Rich feature hierarchies for accurate object detection and semantic segmentation,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.

[28]. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Spatial pyramid pooling in deep convolutional networks for visual recognition,’’IEEETrans.PatternAnal.

Mach. Intell., vol. 37, no. 9, pp. 346–361, Jun. 2014.

[29]. A. Neubeck and L. Van Gool, ‘‘Efficient non-maximum suppression,’’ in Proc. Int. Conf. Pattern Recognit., vol. 3. Aug. 2006, pp. 850–855.

[30]. R. J. Williams, D. E. Rumelhart, and G. E. Hinton, ‘‘Learning representations by back- propagating errors,’’ Nature, vol. 2, pp. 533–536, Oct. 1986.

[31]. I. Sutskever, J. Martens, G. Dahl, and G. Hinton, ‘‘On the importance of initialization and momentum in deep learning,’’ in Proc. 30th Int. Conf. Mach. Learn. (ICML), vol. 3, Feb. 2013, pp. 2176–2184.

[32]. R. Girshick, ‘‘Fast R-CNN,’’ in Proc. ICCV, Dec. 2015, pp. 1440–1448.

[33]. D. Cheng, G. Meng, G. Cheng, and C. Pan, ‘‘SeNet: Structured edge network for sea–land segmentation,’’ IEEE Geosci. Remote Sens. Lett., vol. 14, no. 2, pp. 247–251, Feb. 2017.

[34]. Z.Liu,L.Yuan,L.Weng,andY.Yang,‘‘Ahi ghresolutionopticalsatellite image dataset for ship recognition and some new baselines,’’ in Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods, 2017, pp. 324–331.

[35]. G.Xia,X.Bai,J.Ding,Z.Zhu,S.Belongie,J. Luo,M.Datcu,M.Pelillo, and L. Zhang, ‘‘DOTA: A large-scale dataset for object detection in aerial images,’’ in Proc. IEEE Comput. Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3974–3983.

[36]. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, ‘‘Image quality assessment: From error visibility to structural similarity,’’ IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[37]. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, ‘‘Improving neural networks by preventing coadaptation of feature detectors,’’ Jul. 2012, arXiv:1207.0580. [Online]. Available: https://arxiv.org/abs/1207.0580

[38]. Z. Ben Xu, Z. Hai, W. Yao, X. Y. Chang, and L. Yong, ‘‘L1/2 regularization,’’ Sci. China Inf. Sci., vol. 53, no. 6, pp. 1159–1169, Jun. 2010.

[39]. Y.Xie, H.W.Kim, H.J.Kim, and T.L.Song, ‘‘Reduction of computational load for implementing iJIPDA filter,’’ in Proc. 20th Int. Conf. Inf. Fusion, Jul. 2017, pp. 1–6.