Using convolutional neural network (CNN) approach for ship detection in Sentinel-1 SAR imagery

Hashir Tanveer
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS)
Wuhan University
Wuhan, China
hashir@whu.edu.cn

Timo Balz
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS)
Wuhan University
Wuhan, China
balz@whu.edu.cn

Bahaa Mohamdi
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS)
Wuhan University
Wuhan, China
bh.mo@whu.edu.cn

Accurate maritime ship surveillance and monitoring ensures compliance with port regulations and standards. The growing volume of waterborne traffic however, has made this goal difficult to achieve in applications like maritime traffic control, ship search and rescue, territorial regulation, and fishery management. Detection of ships is complicated, especially under unfavourable conditions, such as during night-time or on cloudy days. Synthetic aperture radar (SAR) provides high-resolution data that can overcome these limitations. Using machine-learning techniques to detect ships in a SAR based image can increase the accuracy of identification detection results as compared to traditional image-based object detection methods. Sentinel-1 SAR images from 2015 to 2018 were used in this exploratory study presenting an analysis of effective ship detection and ship count in a congested sea environment using a Convolutional Neural Networks (CNNs) method, the Faster R-CNN VGG16. Experiments using sixteen convolutional layers in the model yielded promising and significant detection, identification and ship count results for the port of Shanghai.

Keywords— Ship Detection, SAR, Machine Learning, Faster-RCNN (VGG16)

I. Introduction

Satellite images provide a large ground coverage extending across water covered areas, making possible continuous observation of ship locations. Ship detection has been active research topic in the field of remote sensing for a long time and supports maritime rescue, cargo transportation, and port management. Many ship detection methods have been proposed, however this still constitutes a challenge due to uncertainties such cloud cover, disruptors, ship density and so on. Satellite radar data has shown great value for monitoring maritime activities. Synthetic Aperture Radar (SAR) is especially useful for detection and monitoring the position of vessels over time [1][2]. With the launch of the Sentinel-1A and Sentinel-1B SAR satellites, the availability of radar data has significantly improved. This satellite mission continuously produces six terabytes of data everyday [3]. Vessel and ship detection using SAR is already operational, while classification of ships remains a problem to be tackled.

In recent years, many traditional ship detection methods have been proposed [4-8]. Some methods use the idea of sea-land segmentation to extract sea regions as regions of interest, while algorithms like the contrast box algorithm [9] and semi-supervised hierarchical classification [10] are applied to acquire an object region. False box filtering during post processing yields object region results. A bottom-up visual attention mechanism can select prominent candidate regions throughout a detection scene [11]. Normally, the detection delivers satisfactory results under normal conditions where ships are at certain distance from each other. Even though these methods have produced significant results, they have impracticality in complicated scenarios because of ambiguities such as the density of ships in a small area of interest. This situation is very common at seaport areas because of incoming and outgoing shipping.

In recent years, tremendous progress in object detection and classification has been made using Convolutional Neural Network (CNN) based methods that can detect and extract features automatically. Some CNN models like Alexnet [12], GoogleNet [13], and Resnet [14] are capable of handling large datasets at high accuracy, with excellent potential for object classification and detection. Due to high accuracy in object classification and recognition in ocean surveillance, much research has engaged with CNN for marine ship classification and recognition.

With advancement in neural convolution network and deep learning applications in object detection, many effective algorithms have been proposed [15-21]. For example, region proposals using CNNs (RCNN) [22], Spatial Pyramid Pooling Network (SSP-Net) [23], and Fast-RCNN [24]. Faster-RCNN [25] uses Region Proposal Network (RPN) structure and enhance the detection efficiency while attains end-to-end training. Rather than relying on regional proposals, You Only Look Once (YOLO) [26] and Single Shot MultiBox Detector (SSD) [27] directly approximate the object region and enable real-time detection. Feature Pyramid Network (FPN) [28] uses a multi-scale feature pyramid and makes full use of the feature map in order to obtain accurate detection results. Region-based Fully Convolutional Networks (R-FCN) [29] generates a fully convolution network, which highly reduces the number of parameters, enhances the detection speed, and has a better detection effect.

Using machine-learning techniques to detect ships in a SAR based image can improve the accuracy of results. Models that use machine learning are a hundred times faster than traditional neural network. The advanced framework for detection delivers results that are more reliable, thus ensuring accuracy and efficient utilization of information.

II. Methodology

Conventional ship detection methods comprise the following main stages: land masking, pre-processing, and discrimination. This study explores how the faster R-CNN

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

VGG16 neural network technique can be applied for ship detection and counting in congested sea areas, using SAR images. We used an automatic object detection model to accurately detect and recognize ships using feature details. In the first stage, pre-processing of the sentinel-1 images is carried out. In pre-processing, back-geocoding, averaging, binary conversion, and image morphology are performed. After image morphology, the reflectance values of the images are normalized to set a standard reflectance value. The convolutional neural layers are used to classify the ships in SAR images. The process is shown in the figure 1 below.

Fig 1. Process Diagram

Figure 1 illustrates the major process of ship detection and count. This methodology is designed to detect and categorize the ship in congested areas of seaports where ships are close to each other.

A. Pre-processing of SAR images

Sentinel-1A SAR images used in this study are single look complex images. Sub swaths of forty images are processed and back-geocoded. Back-geocoding is a digital elevational model (DEM) based coarse registration of images. A single back-coded images is shown in figure 2.

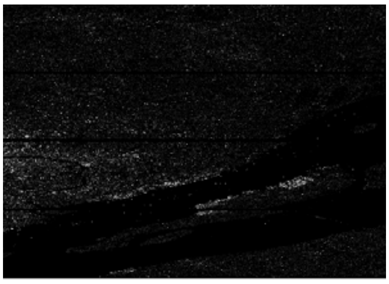

Fig 2. Single back-coded SAR image

A median of all these back-geocoded images is generated using this following formula.

$$\{(n + 1) \div 2\}\text{th} \quad (1)$$

where "n" is the number of items in the set and "th" just means the (n)th number.

Median image is shown in figure 3.

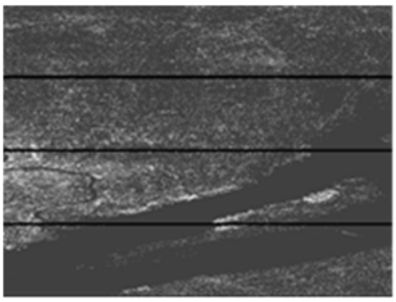

Fig 3. Median image

Single back-geocoded images are differenced with median image in order to get a final image using following formula.

$$\text{Differenced image} = \text{single image} - \text{median image} \quad (2)$$

A final differenced image is shown in fig 4.

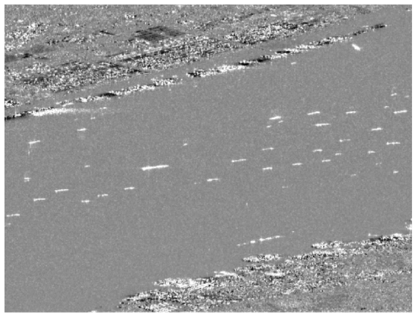

Fig 4. Differenced image

Each differenced image is then converted into binary. The image is then refined and performed adaptive adjustment to achieve the desired bright points on the black background by setting a threshold. A binary image is shown in figure 5.

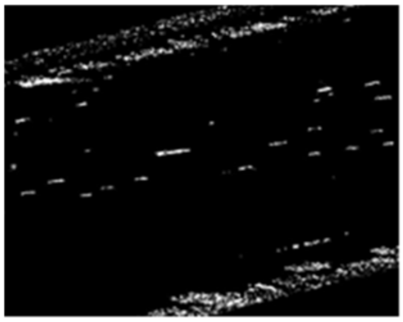

Fig 5. Binary image with specific threshold

The elements on the surface are not very clear, because there are many holes in it. In order to make the elements clearer and visible, a process of "dilation" is carried out. This process is a mathematical morphology operation that is used to enhance the pixel size in a binary image. Figure 6 illustrates the process of dilation.

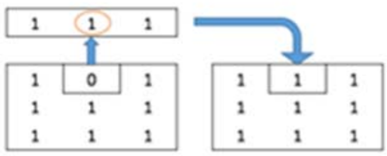

Fig 6. Dilation in a binary image

This process uses the highest value of all the pixels in the neighbourhood as the value of the corresponding pixel in the output image. After that, the dilated image is further processed and a threshold is set to remove small elements in the image. The dilated image with threshold is shown below in fig 7.

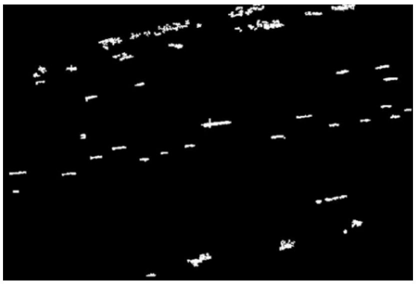

Fig 7. Dilated image with removed small objects.

In figure 7 all the small holes are filled after the successful implementation of dilation process. This process will help in better identification and extraction of ship feature that will lead to the accurate detection of ship. After the process of dilation, the normalized value of pixel intensities in images are calculated to set a standard threshold. The normalized pixel intensity values are shown in the figure 8 below.

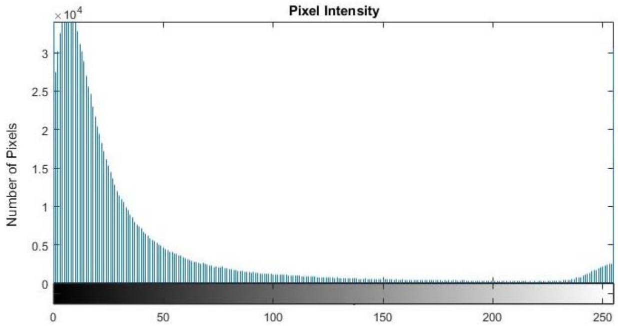

Fig 8. Average pixel intensity values

Figure 8 shows the excessive in concentration of black pixels however, we see number of bright intensities on the right side which depicts the brighter pixels of ships in the image. This pixel intensity is used to set a standard threshold for discriminating ships from other bright (grey or light-grey) pixels with less intensity.

B. *Target classification model based on convolutional neural network (CNN)*

Target classification of an object can be defined as a transformation of a feature space into nominal set values. In order to detect ships, we have used two faster R-CNN models. In this process of classification, an image passes through a series of convolutional network layers, non-linear, pooling, and fully connected layer to get an output.

In the architecture of faster R-CNN, the process starts with images from which we need bounding boxes and the label of those boxes. The input images are passed through a previously trained CNN model and end up in a convolutional feature map. Then region proposal network (RPN) that uses features computed by CNN to locate and obtain a pre-defined number of regions (bounding boxes) that may have objects. Region of interest (ROI) pooling corresponds to the relevant objects extracted by CNN. Finally, R-CNN modules classify the element in bounding boxes and adjust their coordinates. Ships are categorized into three sizes, small, medium and big ships. The faster R-CNN architecture is shown in fig 8.

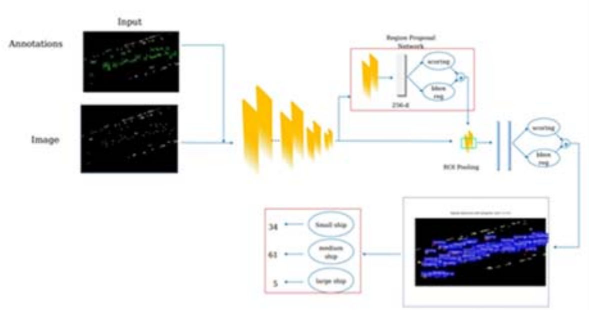

Fig 8. Ship Detection CNN architecture

C. *Data Specification*

The study area used in this research is port of Shanghai in the east of China. The data used in this research study is Sentinel -1 SAR imagery with VH polarization taken from May 15, 2016 to September 26, 2018.

D. *Ship detection*

Ships have higher radar cross-section (RCS) than the adjacent sea clutter in SAR images. It is because of the double bounce effect of incoming radar waves from the ship's structure. So, by setting an appropriate threshold, ships can be detected and separated from the sea clutter. In the process of ship detection, image pixels are classified into two classes of ship and sea. As in the pre-processing step, radiometric correction is already done. Then the image is further processed in the convolutional network for extraction, selection, and detection. In ship detection, image pixels are classified into two categories of ship and water. Fig 9 shows a flowchart of the detection process.

Fig 9. Ship detection process

In the stage of pre-processing back-geocoding, averaging of images, image differencing, thresholding and dilation. Ship features are then extracted. After that the extracted ship features are annotated into neural network layers. Finally, ships are detected with two methods of convolutional neural networks.

III. RESULTS

Images are cropped into a smaller size to apply neural network process to detect ships. Results of vgg16 show promising output. Ships are distinguished and detected using training data model (ground truth). A minor variation in detection of ships is found in between these two techniques. Faster R-CNN Vgg16 took almost 24 hours to complete the process with 30 images. Results of detection are separated with two colors. The yellow box represents a ground truth while the red box indicates the predicted ship in the image. Detection results of vgg16is shown in figure 10.

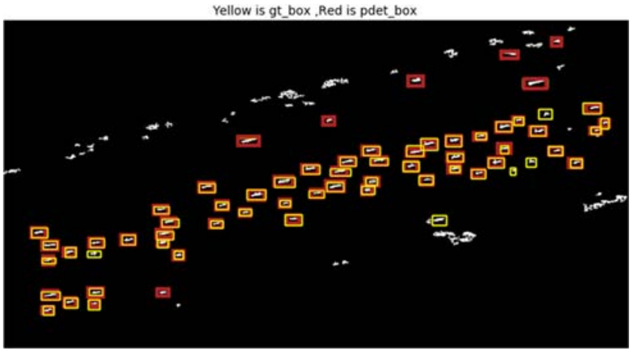

Figure 10. Ship detection with Vgg16

A. Discussion

The difference in the results of ground truth objects and predicted objects is minor. The number of ships counted by faster R-CNN vgg16 is a little bit more than the number of ships selected in ground truth data. In vgg16 the model also detected some of the containers on the port as ship due to the similarity in shape and texture. Furthermore, some of the ground truth training data are not predicted as the ship. Number of ships counted by VGG16 in each image is shown as a graph in fig 10.

Fig 10. Number of ships

In this process, some of the ships with very small-scale gap were also detected as separated ship and counted with as two ships. An average of 69 ships are detected per image with a significant number of more ships in months of May and December.

IV. Conclusion

This paper used convolutional neural network (CNN) model named Faster R-CNN VGG16 to detect and count ships. This method can guarantee highly satisfactory estimation of ships. In contrast to Faster R-CNN VGG16, other CNN could utilize relatively more time while processing due to the complexity and difference in the number of convolutional layers. Detection and count of ships are very promising in this technique with better accuracy, especially in the water. However, the accuracy is slightly lower on the harbor due to many metal containers. Furthermore, in order to save time Faster R-CNN VGG16 might be a better option which can be efficient by feeding more training data to the model.

References

[1] Kourti, N.; Shepherd, I.; Greidanus, H.; Alvarez, M.; Aresu, E.; Bauna, T.; Chesworth, J.; Lemoine, G.; Schwartz, G. Integrating remote sensing in fisheries control. Fisheries Manag. Ecol. 2005, 12, 295–307. [CrossRef]

[2] Zhao, Z.; Ji, K.; Xing, X.; Zou, H.; Zhou, S. Ship surveillance by integration of space-borne SAR and AIS—Review of current research. J. Navig. 2014, 67, 177–189.

[3] Wagner, W. Big Data infrastructures for processing Sentinel data. Photogramm. Week 2015, 15, 93–104.

[4] Crisp, D.J. A ship detection system for RADARSAT-2 dual-pol multi-look imagery implemented in the ADSS. In Proceedings of the 2013 IEEE International Conference on Radar, Adelaide, Australia, 9–12 September 2013; pp. 318–323.

[5] Crisp, D.J. The state-of-the-art in ship detection in Synthetic Aperture Radar imagery. Org. Lett. 2004, 35, 2165–2168.

[6] Wang, C.; Bi, F.; Zhang,W.; Chen, L. An Intensity-Space Domain CFAR Method for Ship Detection in HR SAR Images. IEEE Geosci. Remote Sens. Lett. 2017, 14, 529–533.

[7] Fingas, M.F.; Brown, C.E. Review of Ship Detection from Airborne Platforms. Can. J. Remote Sens. 2001, 27, 379–385.

[8] Leng, X.; Ji, K.; Zhou, S.; Zou, H. An adaptive ship detection scheme for spaceborne SAR imagery. Sensors 2016, 16, 1345.

[9] Yu, Y.D.; Yang, X.B.; Xiao, S.J.; Lin, J.L. Automated Ship Detection from Optical Remote Sensing Images [J]. Key Engineering Materials. IEEE Geoscience & Remote Sensing Letters. 2012, 9, 749-753.

[10] Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features. IEEE Transactions on Geoscience & Remote Sensing. 2010, 48, 3446-3456.

[11] Bi, F.; Zhu, B.; Gao, L.; Bian, M. A Visual Search Inspired Computational Model for Ship Detection in Optical Satellite Images. IEEE Geoscience & Remote Sensing Letters. 2012, 9, 749-754.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions, 2014.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[15] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Seattle, WA, USA, 27–30 June 2016; pp. 2818-2826.

[16] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D., et al. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 1-9.

[17] APASimonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. Computer Science. 2014.

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

[19] Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning. 2015; pp. 448-456.

[20] Simonyan, K.; Zisserman, Very deep convolutional networks for large-scale image recognition. Computer Science. 2014.

[21] He, K.; Sun, J. Convolutional neural networks at constrained time cost. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), Columbus, OH, USA, 23–28 June 2014; pp. 5353-5360.

[22] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

[23] He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2015, 37, 1904-1916.

[24] Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 1440-1448.

[25] 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.

[26] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 779–788.

[27] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Fu, C.; Berg, A.C. SSD: Single Shot Multibox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, Netherlands, 8–16 October 2016; pp. 21–37.

[28] Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. arXiv 2016 arXiv:1612.03144.

[29] Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems. 2016; pp. 379-387.

[30]