Airbus Ship Detection from Satellite Imagery using Frequency Domain Learning

Mina Al-Saad[a], Nour Aburaed[a,c], Alavikunhu Panthakkan[a], Saeed Al Mansoori[b], Hussain Al Ahmad[a], and Stephen Marshall[c]

[a]University of Dubai, Dubai, UAE
[b]Mohammed Bin Rashid Space Centre, Dubai, UAE
[c]University of Strathclyde, Glasgow, UK

ABSTRACT

Ship detection from remote sensing images has been a topic of interest that gradually gained attention over the years due to the wide variety of its applications in the field of maritime surveillance, such as oil discharge control, sea pollution monitoring, and harbour management. Even though there is an extensive amount of methods developed for ship detection, there are still several challenges that remain unsolved, especially in complex environments. These challenges include occlusions due to shadows, clouds, and fog. Nowadays, deep learning algorithms, especially Deep Convolutional Neural Networks (DCNNs), are considered as a powerful approach for automatically detecting ships in satellite imagery. In this paper, enhanced Faster R-CNN (FRCNN) model will be used to overcome the aforementioned unsolved challenges. The enhanced FRCNN, which combines high level features with low level features, will be trained and tested in the frequency domain using the publicly available satellite imagery dataset, Airbus Ship Detection, provided by Kaggle. The performance will be compared to the original FRCNN based on their Overall Accuracy (OA) and Mean Average Precision (mAP) metrics.

Keywords: Deep Learning, FRCNN, Discrete Wavelet Transform, Remote Sensing, Ship Detection

1. INTRODUCTION

Since 1978, ship detection from remote sensing imagery has been a significant image processing task for a wide range of applications, such as maritime traffic flow monitoring, illegal fisheries management, and surveillance of sewage from ships.[1] With the rapid development of remote sensing technology and the availability of high-resolution imagery, it became possible to capture small objects, such as ships, from satellite imagery. The traditional methods to detect ships from satellite imagery were considered both time consuming and labour intensive. Therefore, it is essential to develop algorithms that automatically detect ships from remote sensing imagery.[2] Various factors affect the detection of ships from the optical remote sensing imagery, such as viewpoint changes, cloud cover, and background clutter. Also, the characteristics of optical remote sensing images, such as the diversity of target size and background complexity, render the ship identification and detection task particularly challenging.[3] Nowadays, Deep Convolutional Neural Networks (DCNN) are utilized to accomplish several image processing tasks, such as semantic segmentation[4],[5] object detection[6] and super-resolution[7].[8] Many research studies have been focused on utilizing Deep Learning (DL) techniques, especially DCNN, which have the advantage of being able to automatically extract reliable features that facilitate ship detection. Such approaches showed remarkable achievement for tackling ship detection problems.[9–12] On the contrary to the traditional target detection algorithms, DCNN-based detection algorithms ease the process of detecting ships with a variety of different sizes, shapes, and colours from high-resolution satellite imagery.[13] Current state-of-the-art object detection approaches based on DL include Region-based Convolutional Neural Network (RCNN) series,[14] which have been widely applied in the field of ship detection. RCNN was later upgraded to Fast RCNN, which in turn was upgraded to Faster RCNN (FRCNN). In FRCNN, a Region Proposal Network (RPN) is introduced to obtain

Further author information: (Send correspondence to M.A.S.)
M.A.S.: E-mail: malsaad@ud.ac.ae,
N.A.: E-mail: noaburaed@ud.ac.ae

An peer reviewed accepted author manuscript of the following research article: Al-Saad, M., Aburaed, N., Panthakkan, A., Al Mansoori, S., Ahmad, H. A., & Marshall, S. (2021). Airbus ship detection from satellite imagery using frequency domain learning. Proceedings of SPIE, 11862, [1186210]. https://doi.org/10.1117/12.2600168

the proposal regions from the feature maps through Region of Interest (ROI) pooling and that can significantly reduce the proposal cost compared to the popular Selective Search method.[15] Zhang et al.[13] developed an algorithm for detecting and identifying small ships from high-resolution remote sensing imagery based on a modified version of FRCNN. The authors use VGG-19 as a backbone for FRCNN and improve it by using multiresolution convolutional features. Additionally, segmentation of water and non-water areas is performed before the extraction of the candidate ROI that may contain a ship. This approach performs better than the original FRCNN, but the accuracy is dependent on the accuracy of water and non-water segmentation. Other works[16] proposed an improved FRCNN architecture by introducing dilated convolution, which improves the capability of feature extraction. Dark channel prior method was used as a pre-processing step to dehaze input images, and then Gray World algorithm was applied to balance the color constancy of the dehazed image. The performance of the proposed methodology was evaluated using HRSC2016 dataset, which shows superior results compared to the original FRCNN model. Another modified version of FRCNN was proposed in[17] for small object detection in optical remote sensing images. The model not only uses ROI from RPN, but also incorporates context information to further boost the detection of small objects. The problem of imbalanced class distribution during the training is also solved by introducing an effective data augmentation method known as Random Rotation during the training process. The goal of this research paper is to enhance the performance of FRCNN by incorporating wavelet decomposition into its input architecture. Rather than training the network using RGB input directly, each input band is decomposed into its wavelet components, which are then merged together and used for training. The goal is to achieve higher accuracy with an input size lower than the original image by utilizing wavelet domain. The rest of the paper is organized as follows: section 2 showcases the dataset and pre-processing steps, section 3 discusses the methodology and gives brief background about Discrete Wavelet Transformation (DWT) and FRCNN along with the training and testing procedures, section 4 illustrates the results of the algorithm and compares them against the original FRCNN, and finally, Section 5 summarizes the paper and draws the future direction of this research study.

2. DATASET AND PRE-PROCESSING

The dataset used in this paper is the Airbus Ship Detection Challenge provided by Kaggle.[18] This dataset consists of RGB images of size 768×768. It also includes the encoded pixels that represent ships' locations in the satellite images. The encoded pixels were converted to binary masks, where "ship" is represented as 1 and "no ship" is represented as 0. If a mask of value 1 exists, then it is converted to a bounding box by calculating the four corner coordinates of the mask. In order to preserve computation power, all images were resized to 256×256. The x and y coordinates were flipped as the original data has the axes flipped. An example of an image with one ship bounding box is shown in 1. As this dataset is too large, only images that contain at least one ship were used, so that the final size of the dataset is 18,392 images. However, due to processing limitations and memory restrictions, only 7285 images were selected to train the model with a total of 9000 bounding box coordinates, while 1612 images were used to test and evaluate the performance of the model with a total of 2000 bounding box coordinates.

3. METHODOLOGY

The following subsections describe the architecture of FRCNN, Discrete Wavelet Transform (DWT), and the proposed training strategy for FRCNN in the frequency domain by integrating the wavelet features into the model.

3.1 Faster R-CNN

RCNN[19] was first introduced in 2015 as an answer to PASCAL VOC Challenge. It is an object detector that consists of two stages. The first one is region proposals stage, which extracts ROIs using algorithms such as Selective Search. The second stage is a CNN that takes the ROIs and extracts their features, then classifies them accordingly. Despite the reliability of RCNN, the process of going from the first to the second stage represents a bottleneck. Fast RCNN[20,21] attempts to overcome RCNN drawbacks, where ROI pooling layer is introduced. Nonetheless, Fast RCNN still requires using Selective Search to find ROIs, so bottlenecks impose a problem for this network as well. FRCNN[22] solves this problem by introducing RPN. Similar to its predecessors, FRCNN is

(a) Original RGB image (b) Bounding box mask

Figure 1. Samples from Airbus Ship detection dataset.

a two-stage object-detector; the first model is the RPN, and the second one is the CNN that detects objects from the generated regions.[23] In that sense, FRCNN is considered as an upgraded version of RCNN.[24] The purpose of utilizing RPN is to avoid the slower Selective Search algorithm to select regions of interest and, thus, it speeds up the training process and boosts the overall performance of the model.[25]

3.2 Discrete Wavelet Transform (DWT)

DWT is a mathematical function that was devised in the 1980s, and it is widely used in digital signal processing generally. In image processing, DWT is especially useful for compression and de-noising tasks. Wavelets came as an answer to Fourier Transform (FT) shortcoming. FT analyzes images in frequency domain but loses time information, thus, abrupt changes cannot be represented efficiently. Wavelets on the other hand are capable of time-frequency analysis. In DWT, the image is decomposed by passing it through high-pass (H) and low-pass (L) filters to extract edges and approximations, respectively. The image is passed through these two filters in different combinations to produce four different components; LL, HL, LH, and HH, as illustrated in Figure 2. LL component represents the approximation of the original image, LH represents the horizontal features, HL represents the vertical features, and HH represents the diagonal features. An image can be decomposed beyond this level by passing LL component through DWT again. The newly produced LL component from the 2nd level can be decomposed further, and so on. There are several families of wavelet methods, for the purpose of this study, Haar DWT[26] is used with only the first level of decomposition.

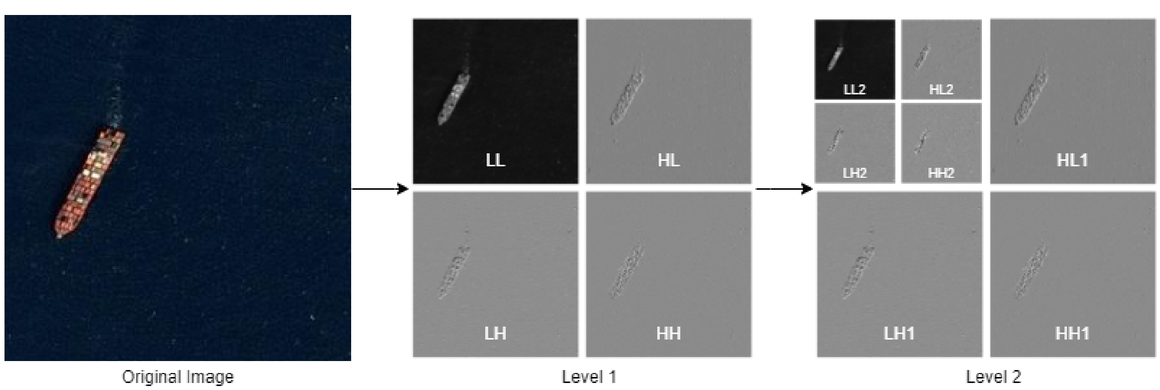

Figure 2. Two wavelet levels using Haar wavelet. Only the first level is considered for this study.

3.3 Proposed Faster R-CNN

The idea of embedding the wavelet transform into the original FRCNN is achieved by adding the wavelet components of the image as part of the data augmentation procedure. It is implemented by decomposing the original image into its high and low frequency components before extracting ROIs, as shown in 3. This procedure increases the amount of the training data and provides extra distinguishable features, which improve the overall learning process. Since the dataset consists of RGB images, each image is split into its 3 individual bands, then each band is decomposed into its four wavelet components. The low and high components obtained are then used for ROI extraction and network training. For instance, the red channel produces the four components LL_r, LH_r, HL_r, and HH_r. Green and blue produce their own four components as well. Afterwards, several combinations of low and high components fusion are tested to see which combination yields the best performance. The combinations tested are case 1:(LL_r, LL_g, LL_b), case 2:(LL_g, LH_r, HL_b), and case 3:(LL_r, LH_g, HL_b). Each combination is tested under the same conditions along with the original version of the network without DWT in order to provide fair comparison. The network was trained over 50 epoch using Adam optimizier with Resnet-50 backbone. The training parameters are summarizes in Table 1.

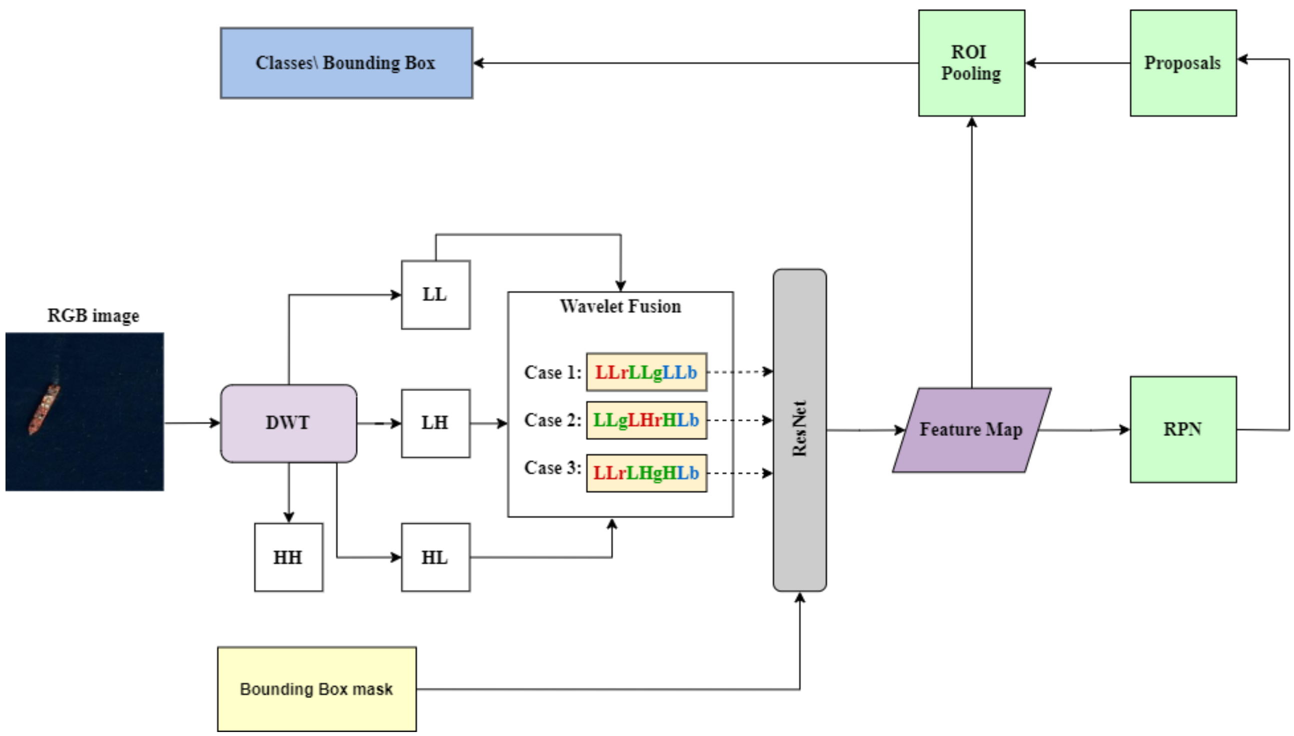

Figure 3. An overview of the proposed FRCNN. The dashed lines in the diagram convey that each wavelet combination is considered as separate cases, and each case is trained individually.

4. RESULTS

The enhanced FRCNN and the original one were both developed and trained using Python Keras (2.3.1) library with Tensorflow backend using NVIDIA Quadro P6000-24GB GPU and Intel(R) 12 core Processor CPU with 380GB RAM. The results of training and testing the networks were evaluated in terms of the Overall Accuracy (OA) and Mean Average Precision (mAP). OA is calculated as follows:

$$OA = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

Table 1. Wavelet FRCNN Training Parameters.

Parameter	Value
Epochs	50
Optimization Function	Adam
Backbone	Resnet-50
Learning rate	0.00001
Batch size	7

Where TP stands for True Positive, TN stands for True negative, FP stands fore False Positive, and FN stands for False Negative. In object detection, the class label is not enough to decide whether the prediction is TP or otherwise, as the bounding box needs to be taken into consideration as well. For each bounding box, the Intersection over Union (IoU) is measured as follows:

$$IoU = \frac{Area\,of\,Overlap}{Area\,of\,Union} \tag{2}$$

IoU is given a threshold value to decide the outcome. For instance, if the IoU threshold is 0.5 and the IoU value for a prediction is 0.8, then the prediction is considered TP. However, if the predicted IoU value was less than 0.5, the prediction is considered FP. For this study, the IoU threshold is 0.5. mAP is defined as the area under the Precision-Recall curve, where Precision is defined as:

$$P = \frac{TP}{TP + FP}, \tag{3}$$

and Recall is defined as:

$$R = \frac{TP}{TP + FN}. \tag{4}$$

Finally, AP can be described with the following equation:

$$AP = \int_0^1 P(R), \tag{5}$$

mAP is calculated as the average of AP for each class. In this case, since the purpose is to detect ships only, AP and mAP are considered equal. The results of all three cases are summarized in Table 2 and compared to the original FRCNN. As seen in the table, Case 1, which is composed of $LL_r LL_g LL_b$ yielded the highest accuracy and mAP compared to the original and two other cases. The results of Case 2 are also higher than the original FRCNN. This asserts the fact that training with wavelets boosts the performance with an input of a lower size than the original one.

Table 2. Results summary of wavelet FRCNN compared to the original one.

Evaluation metrics	Wavelet FRCNN			Original FRCNN
	Case 1	Case 2	Case 3	
OA	**93.71%**	93.15%	92.16%	92.70%
mAP	0.82 ± 0.03	0.76 ± 0.02	0.77 ± 0.02	0.80 ± 0.03

5. CONCLUSION

This paper introduces enhancements to FRCNN performance in terms of accuracy and mAP by utilizing wavelet decomposition to train the network in the frequency domain. The dataset taken from Airbus Ship Detection Challenge on Kaggle was pre-processed and decomposed into its high and low frequency components. Through testing various combinations of these components, it has been proven that training the network with $LL_rLL_gLL_b$ combination achieves higher performance than the original one. Some other combinations achieved a higher accuracy as well, such as $LL_gLH_rHL_b$. For the future direction of this research, wavelet enhancements will be tested on other state-of-the-art object detection networks, such as You Only Look Once (YOLO) series, Single Shot Multibox Detection (SSD), and RetinaNet.

REFERENCES

[1] Xie, X., Li, B., and Wei, X., "Ship detection in multispectral satellite images under complex environment," *Remote Sensing* **12**(5), 1–19 (2020).

[2] Wang, R., You, Y., Zhang, Y., Zhou, W., and Liu, J., "Ship detection in foggy remote sensing image via scene classification r-cnn," in [*2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*], 81–85, IEEE (2018).

[3] Wu, Y., Ma, W., Gongand Z. Baiand W. Zhao, M., Guo, Q., Chen, X., and Miao, Q., "A coarse-to-fine network for ship detection in optical remote sensing images," *Remote Sensing* **12**(2), 1–24 (2020).

[4] Talal, M., Panthakkan, A., Mukhtar, H., Mansoor, W., Almansoori, S., and Al Ahmad, H., "Detection of water-bodies using semantic segmentation," in [*2018 International Conference on Signal Processing and Information Security (ICSPIS)*], 1–4, IEEE (2018).

[5] El Rai, M. C., Aburaed, N., Al-Saad, M., Al-Ahmad, H., Al Mansoori, S., and Marshall, S., "Integrating deep learning with active contour models in remote sensing image segmentation," in [*2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*], 1–4, IEEE (2020).

[6] Al Mansoori, S., Aburaed, N., Al-saad, M., AlShamsi, M., and AlMaazmi, A., "Iaf earth observation symposium (b1)," in [*Automatic Ship Detection from High Resolution Satellite Images Based on a Deep Convolutional Neural Network (DCNN) Model*], Proceedings of the International Astronautical Congress, IAC (2019).

[7] Aburaed, N., Panthakkan, A., Al-Saad, M., El Rai, M. C., Al Mansoori, S., Al-Ahmad, H., and Marshall, S., "Super-resolution of satellite imagery using a wavelet multiscale-based deep convolutional neural network model," in [*Image and Signal Processing for Remote Sensing XXVI*], Bruzzone, L., Bovolo, F., and Santi, E., eds., **11533**, SPIE (2020).

[8] Aburaed, N., Panthakkan, A., Almansoori, S., and Al-Ahmad, H., "Super resolution of ds-2 satellite imagery using deep convolutional neural network," in [*Image and Signal Processing for Remote Sensing XXV*], Bruzzone, L. and Bovolo, F., eds., **11155**, 485–491, International Society for Optics and Photonics, SPIE (2019).

[9] Li, X. and Cai, K., "Method research on ship detection in remote sensing image based on yolo algorithm," in [*2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*], 104–108, IEEE (2020).

[10] Wang, R., F.Xu, Pei, J., Wang, C., Huang, Y., Yang, J., and Wu, J., "An improved faster r-cnn based on mser decision criterion for sar image ship detection in harbor," in [*IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*], 1322–1325 (2019).

[11] Li, Y., Zhang, S., and Wang, W. Q., "A lightweight faster r-cnn for ship detection in sar images," *IEEE Geoscience and Remote Sensing Letters* , 1–5 (2020).

[12] Sali, S. M., Manisha, N. L., King, G., and Vidya Mol, K., "A review on object detection algorithms for ship detection," in [*2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*], **1**, 1–5 (2021).

[13] Zhang, S., Wu, R., Xu, K., Wang, J., and Sun, W., "R-cnn-based ship detection from high resolution remote sensing imagery," *Remote Sensing* **11**(6), 1–15 (2019).

[14] Ren, S., He, K., Girshick, R., and Sun, J., "Ship target detection and identification based on ssd_mobilenetv2," in [*IEEE Transactions on Pattern Analysis and Machine Intelligence*], **39**(6), 1137–1149 (2017).

[15] Liu, B., Zhao, W., and Sun, Q., "Study of object detection based on faster r-cnn," in [*2017 Chinese Automation Congress (CAC)*], 6233–6236, IEEE (2017).

[16] Wei, S., Chen, H., Zhu, X., and Zhang, H., "Ship detection in remote sensing image based on faster r-cnn with dilated convolution," in [*2020 39th Chinese Control Conference (CCC)*], 7148–7153 (2020).

[17] Ren, Y., Zhu, C., and Xiao, S., "Small object detection in optical remote sensing images via modified faster r-cnn," *Applied Sciences* **8**(5), 1–11 (2018).

[18] Airbus, "Airbus Ship Detection Challenge." Kaggle, 2019 https://www.kaggle.com/c/airbus-ship-detection. (Accessed: 11 August 2021).

[19] Girshick, R., Donahue, J., Darrell, T., and Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR* **abs/1311.2524** (2013).

[20] Girshick, R. B., "Fast R-CNN," *CoRR* **abs/1504.08083** (2015).

[21] Girshick, R., Donahue, J., Darrell, T., and Malik, J., "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**(1), 142–158 (2016).

[22] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR* **abs/1506.01497** (2015).

[23] Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., and Chao, D., "Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network," *Remote Sensing* **12**(9), 1–25 (2020).

[24] Aburaed, N., Al-Saad, M., El Rai, M. C., Al Mansoori, S., Al-Ahmad, H., and Marshall, S., "Autonomous object detection in satellite images using wfrcnn," in [*2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*], 106–109 (2020).

[25] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," in [*IEEE Transactions on Pattern Analysis and Machine Intelligence*], **39**(6), 1137–1149 (2017).

[26] Haar, A., "Zur theorie der orthogonalen funktionensysteme," *Mathematische Annalen* **69**, 331–371 (Sep 1910).