2019「中技社科技獎學金」

2019 CTCI Foundation Science and Technology Scholarship

境外生生活助學金

Living Grant for International Graduate Students

SAR IMAGE SHIP DETECTION BASED ON YOLOV2 DEEP LEARNING FRAMEWORK

Yang-Lang Chang[1,*], *Amare Anagaw*[1], *Lena Chang*[2] , *Yi Chun Wang*[1],*Chih-Yu Hsiao*[1], *Wei-Hong Lee*[1]
National Taipei University of Technology, Taiwan, [2]National Taiwan Ocean University, Taiwan

Abstract

The remote sensing data are huge in nature which brings a challenge for real time object detection. However, to mitigate this problem a high performance computing (HPC) methods have been proposed to accelerate SAR imagery analysis, especially the GPU based computing methods. In this paper, we propose an enhanced GPU based deep learning method to detect ship from the SAR images. The *you only look once version 2* (YOLOv2) deep learning framework is proposed to model the architecture and training the model. Additionally, in order to reduce computational time with relatively competent detection accuracy, we develop a new architecture with less number of layers called *YOLOv2-Reduce*. In the experiment, we use two types of dataset: SAR Ship Detection Dataset (SSDD) dataset and *Diversified SAR Ship Detection Dataset* (DSSDD). YOLOv2 test results showed an increase in accuracy of ship detection as well as a noticeable reduction in computational time compared to Fast R-CNN. The proposed *YOLO-Reduce* architecture has a competent detection performance as YOLOv2, but with less computational time on NVIDIA TITAN X GPU.

1. Introduction

Ship detection is an important topic in the field of remote sensing. At present, many object detection methods have developed in pattern recognition community. However, many of the proposed systems have computationally intensive problems for high accuracy performance. The traditional methods of target detection are roughly divided into region selections, e.g., SIFT, and HOG, and classifier, e.g., SVM, and Adaboost. After AlexNet won ImageNet's image classification in 2012 with very high accuracy and performance in object detection using deep learning, the neural network is booming. A SAR image data sets are used for the ship detection in the experiments. We use the YOLOv2 deep learning framework, which is well-known in the field of computer vision, as a base to implement vessel detection and adjust the parameters to achieve high accuracy performance in near real-time. In addition, we introduced a new architecture, *YOLOv2-reduced*, which has fewer layers, by removing some convolutional layer.

2. Methodology

A YOLOv2-based end-to-end training convolutional neural network is constructed to detect ships. First, it uses a single neural network to directly predict the bounding box and class probability. The SAR image is divided into *SxS* grids. Each grid cells predicts the *k* bounding boxes, confidence score of bounding boxes, and class probabilities. YOLOv2 shows a lot of improvements than the original YOLO as shown in Fig. 1. One of the improvements is that it introduces the Faster R-CNN anchor concept into the original framework to improve network performance. YOLOv2 uses these previously experienced anchor boxes to predict bounding boxes and improve the accuracy of the center of the bounding box. Compared to the original YOLO, there are 30 layers of YOLOv2 network architecture as shown in Table 1. In YOLOv2, 22 layers are convolutional layers and 5 layers are the max pooling layers. It has a route layer at the 25th and 27th layers. For example, the 27th route is composed of layer 26 and layer 24, that is, the 26th and 24th layers are merged to the next layer. The role of the route layer is to merge layers. The final detection layer is regresses the features extracted from the convolution layer to predict the probability and the bounding box of the ship. Assuming input image resolution is 416 x 416 x 3, then the output of the 30th layer size is 13 x 13 x 30. Finally, it reduced to a 13 x 13 size grid. The output of each cell is 30 (5 x 6), 5 refers the 5 predictive borders for each 13 x 13 grid cell, and 25 (30 minus 5) means that each border outputs 25 values. One of the six numbers is the probability for the ship. The other five numbers are the position and size of the bounding boxes and the confidence of the bounding boxes.

We use a SAR vessel image dataset, as shown in Fig. 2, which each picture has a ship. In the training task, we manually mark the border and label of each image ship objects as the ground truth. As the PASCAL VOC provides a standardized dataset for object detection, the dataset we used follows the same rule to construct the bounding box and label annotations. In evaluation of ship detection, there is a parameter called *intersection of union* (IoU). It's the overlap rate between the predict bounding box and ground truth generated by the model for evaluating the detection accuracy. When IoU exceeds the threshold, then the bounding box is considered to be correct.

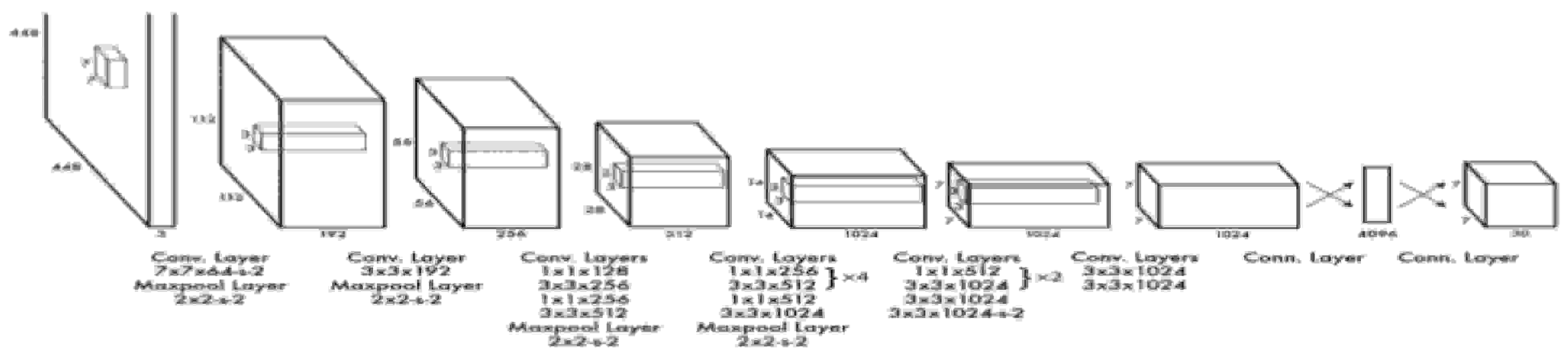

Fig 1. The original YOLO network architecture. Ref. *Joseph Redmon 2016*

3. Experimental Results and Conclusion

We adopt a well-known open source, namely the Darknet framework, to train our deep learning models. Darknet-19 was selected to be the backbone CNN network, which had been pre-trained on VOC 2007+2012. The results of this study verify the correctness and effectiveness of the method by means of accuracy and computational cost. Compared to the faster-R-CNN, the proposed method improves the accuracy to 90.03% on the SSDD dataset as shown in Table 2.

Table. 1 YOLOv2 network architecture

No	Type	Input	Filters	Size/Stride	Output
0	conv	416 x 416 x ___3	__32	3 x 3 / 1	416 x 416 x __32
1	max	416 x 416 x __32		2 x 2 / 2	208 x 208 x __32
2	conv	208 x 208 x __32	__64	3 x 3 / 1	208 x 208 x __64
3	max	208 x 208 x __64		2 x 2 / 2	104 x 104 x __64
4	conv	104 x 104 x __64	_128	3 x 3 / 1	104 x 104 x _128
5	conv	104 x 104 x _128	__64	1 x 1 / 1	104 x 104 x __64
6	conv	104 x 104 x __64	_128	3 x 3 / 1	104 x 104 x _128
7	max	104 x 104 x _128		2 x 2 / 2	_52 x _52 x _128
8	conv	_52 x _52 x _128	_256	3 x 3 / 1	_52 x _52 x _256
9	conv	_52 x _52 x _256	_128	1 x 1 / 1	_52 x _52 x _128
10	conv	_52 x _52 x _128	_256	3 x 3 / 1	_52 x _52 x _256
11	max	_52 x _52 x _256		2 x 2 / 2	_26 x _26 x _256
12	conv	_26 x _26 x _256	_512	3 x 3 / 1	_26 x _26 x _512
13	conv	_26 x _26 x _512	_256	1 x 1 / 1	_26 x _26 x _256
14	conv	_26 x _26 x _256	_512	3 x 3 / 1	_26 x _26 x _512
15	conv	_26 x _26 x _512	_256	1 x 1 / 1	_26 x _26 x _256
16	conv	_26 x _26 x _256	_512	3 x 3 / 1	_26 x _26 x _512
17	max	_26 x _26 x _512		2 x 2 / 2	_13 x _13 x _512
18	conv	_13 x _13 x _512	1024	3 x 3 / 1	_13 x _13 x 1024
19	conv	_13 x _13 x 1024	_512	1 x 1 / 1	_13 x _13 x _512
20	conv	_13 x _13 x _512	1024	3 x 3 / 1	_13 x _13 x 1024
21	conv	_13 x _13 x 1024	_512	1 x 1 / 1	_13 x _13 x _512
22	conv	_13 x _13 x _512	1024	3 x 3 / 1	_13 x _13 x 1024
23	conv	_13 x _13 x 1024	1024	3 x 3 / 1	_13 x _13 x 1024
24	conv	_13 x _13 x 1024	1024	3 x 3 / 1	_13 x _13 x 1024
25	route	16th .			_26 x _26 x _512
26	reorg	_26 x _26 x _512		_ x _ / 1	_13 x _13 x 2048
27	route	26th · 24th .			_13 x _13 x 3072
28	conv	_13 x _13 x 3072	1024	3 x 3 / 1	_13 x _13 x 1024
29	conv	_13 x _13 x 1024	30	1 x 1 / 1	_13 x _13 x __30

Fig 2. Sample from dataset SSDD dataset

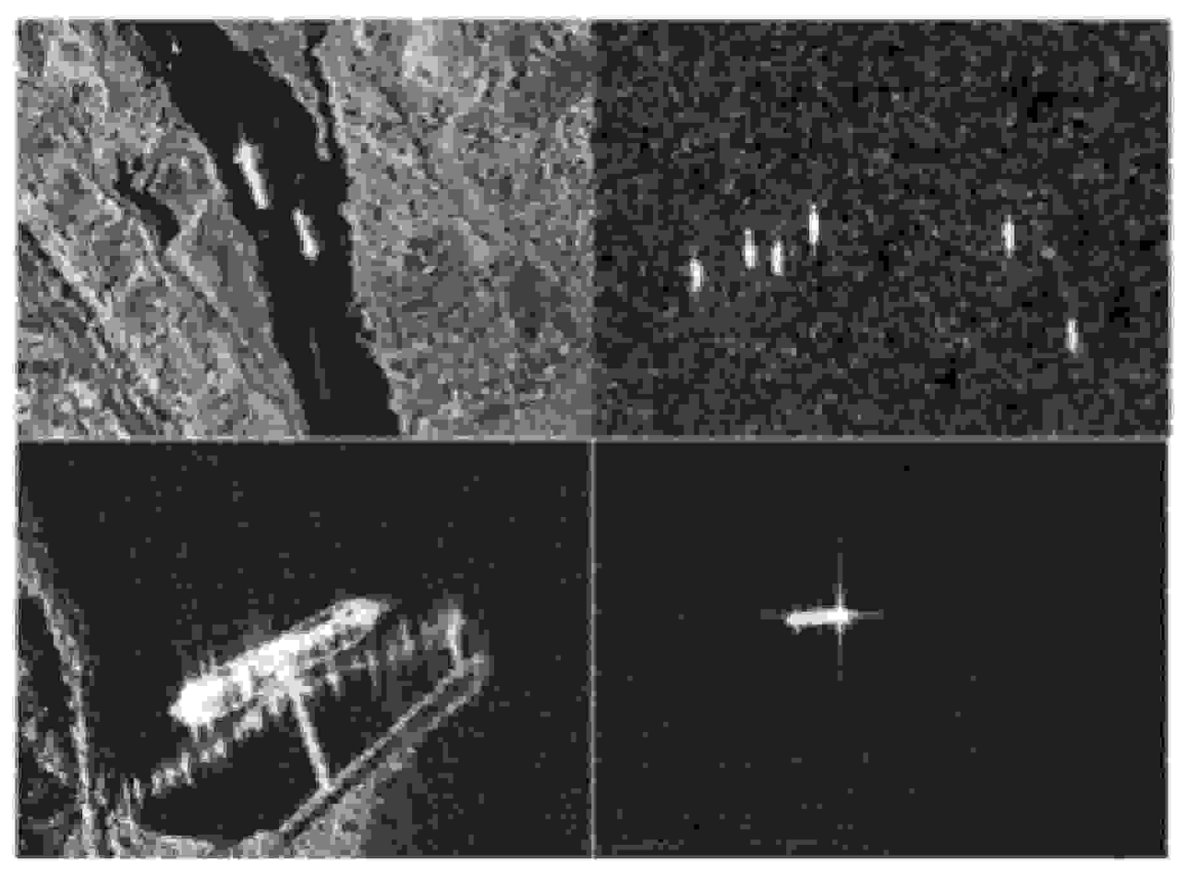

Table. 2 Ship detection accuracy and speed comparison

Networks	Accuracy	Time per image (*ms*)
Faster-R-CNN	70.94 %	206
YOLOv2	90.03 %	35

The newly proposed architecture, YOLOv2-Reduce model, has less number of layers. The repetitive convolution layer is not very effective for ship detection (convolutional layers 23, 24, 25). Since the ship is relatively insignificant to the ocean, applying this consecutive convolution is not required. Therefore, we reduce these three convolutional layers to one layer. This approach reduces the time complexity of YOLOv2 architecture with almost competent detection performance as shown in fig.3.

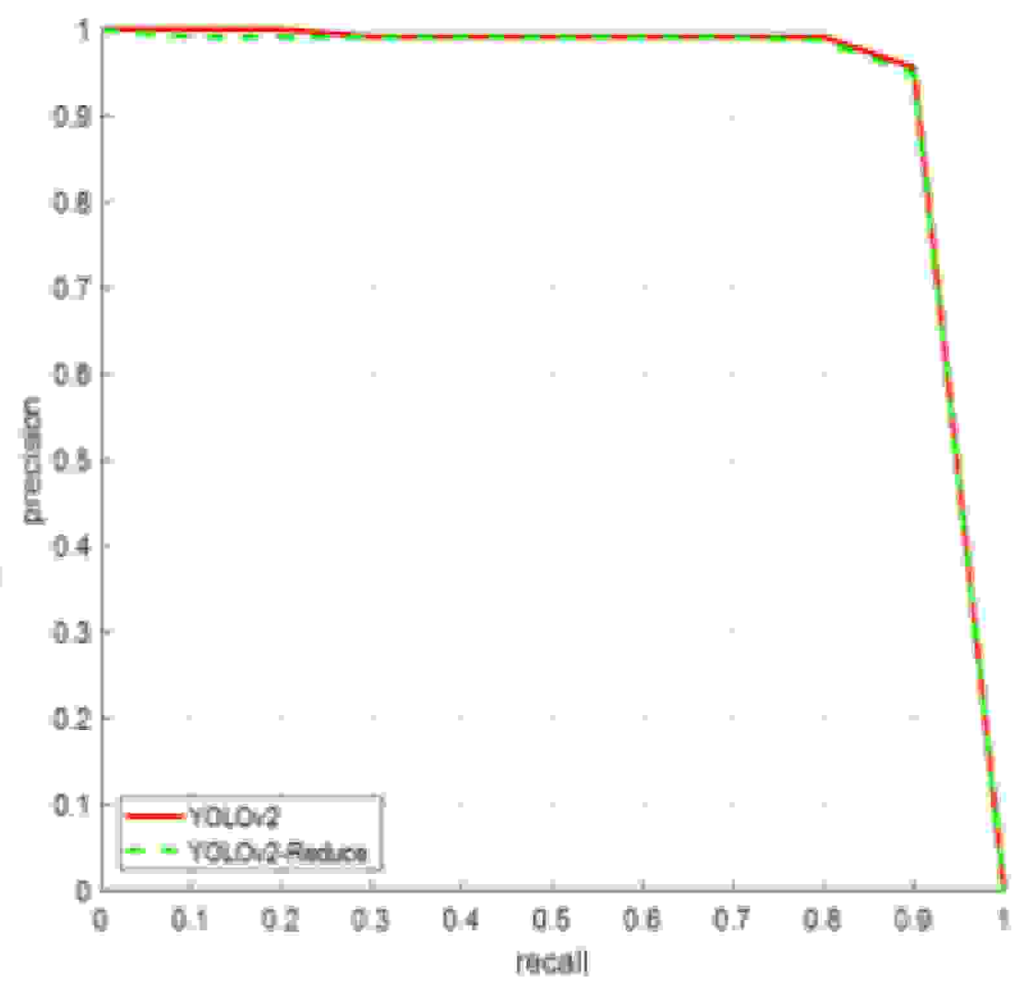

Fig. 3 Precision-Recall curve for YOLOv2 and YOLOv2-reduce on SSDD dataset.

Different cases of the ship detections are shown in Fig. 4. In the faster-R-CNN, the errors often occurred as it was applied to docks, shores or canals. According to this study, YOLOv2 provides not only better accuracy but also more computational performance (5.8x times) than the faster-R-CNN for SAR image ship detection.

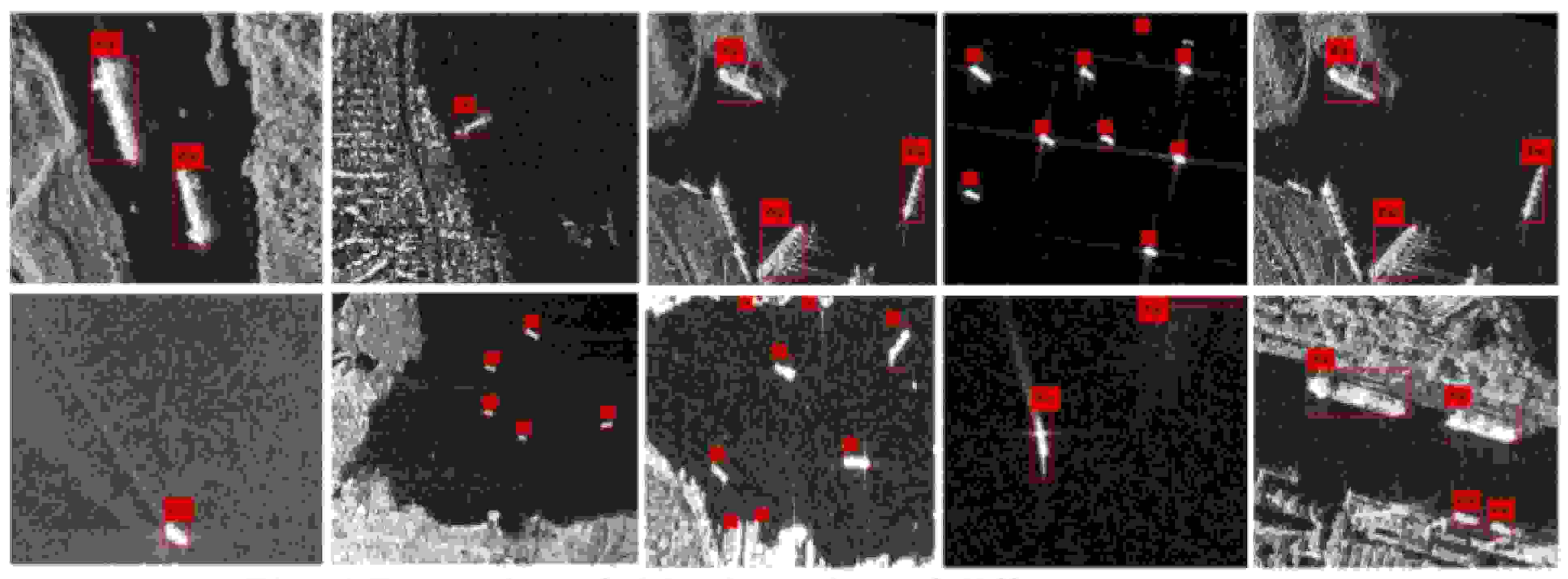

Fig. 4 Examples of ship detection of different cases

PR & EDA Laboratory, Dept. of Electrical Engineering
National Taipei University of Technology, Taipei, Taiwan
E-mail: *ylchang@ntut.edu.tw*
URL: *http://www.ntut.edu.tw/~ylchang/*

TAIPEI TECH

財團法人中技社
CTCI FOUNDATION