A deep learning approach to ship detection using satellite imagery

Marzuraikah Mohd Stofa*, Mohd Asyraf Zulkifley, Siti Zulaikha Muhammad Zaki

Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor, Malaysia

marzuraikah97@gmail.com

Abstract. Automatic ship detection on remote sensing images is one of the important modules in the maritime surveillance system. Its main task is to detect possible pirate threats as early as possible. Thus, the detection system must be accurate enough as it plays a vital role in national security. Therefore, this paper proposes a deep learning approach to detect the presence of a ship in the harbour areas. DenseNet architecture has been selected as the core convolutional neural network-based classifier, where various finetuning has been done to find the optimal setup. The three hyperparameters that have been fine-tuned are optimizer selection, batch size, and learning rate. The experimental results show a success rate of over 99.75% when Adam optimizer is selected with a learning rate of 0.0001. The test was done on the Kaggle Ships dataset with 4,200 images. This algorithm can be further fine-tuned by considering other types of convolutional neural network architecture to increase detection accuracy.

1. Introduction

Automatic ship detection using remote sensing images is very important for maritime security management. Its main functions are to monitor the traffic flow, detect illegal fisheries and prevent marine pollution. For military purposes, the goal of automatic ship detection is to enhance maritime security through Intelligence, Surveillance, and Reconnaissance (ISR) efforts [1]. One of the key technologies that play a crucial role is an advanced remote sensing system, which is used to collect various data. Some of the sensors that have been used in ISR maritime patrol aircraft are Electro-Optical/Infrared (EO/IR) cameras, radar, and Electronic Support Measures (ESM). Since a deep learning approach requires large training data, this paper focuses on satellite images as the data can easily be obtained.

For safety purposes, an Automatic Identification System (AIS) transponder is required to be installed to the commercial ship that exceeds 300 tons and for all passenger ships [2]. This transponder sends data such as ship name, location, destination, and others. However, AIS data can be easily modified. For example, a fisherman boat can impersonate another boat identity by changing types its transmitted information. In addition, recent technologies such as convolutional neural network (CNN), which is a small subset of machine learning (ML) have seen an increasingly successful implementation trend [3]. It also coupled with multi-layered network architecture developed from the classic neural network processes. CNN typically consists of input layers, convolutional layers, activation function, and output layers. It is a learned feature extraction approach and all mappings between input and output are done through the learning process [4].

A deep learning approach has also been used in ship detection and classification. It is inspired by many layers of human brain architecture. After achieving significant success in some artificial intelligence applications such as plantation monitoring [5], disease diagnosis [6], trajectory mapping [7], and physiotherapy [8], a deep learning approach has also been used in Synthetic Aperture Radar (SAR) data processing [9]. Therefore, in this paper, we propose a Dense Convolutional Network (DenseNet) to detect the ship's presence, where it will be considered as a binary classifier problem. DenseNet is one of the convolutional networks that are substantially deeper and accurate because of its feedforward layer. Based on the previous approach, a tracking algorithm is usually employed to
determine the ship’s existence based on specific data and conditions. The same approach can also be applied to ship detection using inverse synthetic aperture radar (ISAR) images and forward-looking infrared (FLIR). Therefore, the main objective of this paper is to obtain high accuracy in the detection and classification of the ship's presence based on satellite images.

2. Recent Works

2.1. Sensor platform used for ship recognition
Kanjir et al. [10] found out that the most widely used sensors in marine surveillance application are optical, infrared and radar sensors. Radar is a typical technology used in ship monitoring and detection which has been used since the 1990s. The satellite-based sensor is also a popular choice for ship detection that requires remote sensing, continuous monitoring, and frequent data collection.

2.2. Object detection approach
There is an increasing research trend in the deep learning-based algorithm for classification purposes. In [11], Lin et al. have introduced a rotational-invariant detection method for object detection in remote sensing images. Their invariant features have delivered good accuracy in the detection of complex objects in remote sensing images. Huang et al. [12] have presented their random forest method, where the training speed is faster compared to the conventional approach but maintaining the same accuracy performance. A novel method based on sparse representation and Hough voting (SR-Hough) was introduced by Yokoya & Iwasaki [13]. This method focuses on detecting instances of an object class or a specific object in the remote sensing images. In [14], Prasad et al. have presented a study that researches on various challenges in maritime surveillance that include occlusion, variations in orientation and scale, multitude number of object classes and changes in weather. A machine learning approach was also introduced by Yu et al. [15] to detect small and dim objects in a Forward Looking Infrared (FLIR) image using a context-driven Bayesian saliency model.

However, these traditional machine learning techniques require explicit object features definition. With recent advancements in deep learning fields and computer vision, learning-based features have drastically changed the conventional approach. The winner of ImageNet Large Scale Visual Recognition Challenges (ILSVRC), Krizhevsky et al. [16] has popularized the usage of deep Convolutional Neural Network (CNN) for image recognition, detection and particularly for classification. Since then, lots of research have been done on CNNs with improvement in various factors such as activation function, optimization technique, network architecture and regularization mechanism [17].

Tang et al. [18] have applied a combined compressed-domain framework, deep neural network and extreme machine learning to detect and classify ships using SPOT-5 dataset images. Zhang et al. [19] also focus on the same goal by proposing the sequential convolutional neural network (S-CNN) method that combines CNN with the saliency detection method. They found out their method performs better compared to the R-CNN. A concise survey of the state-of-the-art saliency object detection method can be found in [20]. Liu et al. [21] also have implemented the CNN approach for ship classification using Google Earth images and reported better results compared to support vector machine and standard neural network. On the other hand, the work in [22] applied the same approach but specializing in the navy application. They introduced a new technique called spatially related detection with convolutional neural network (SPARCNN) that highlights the importance of spatial relationship in improving accuracy.

2.3. Improvement in deep learning
In this modern era, the volume of stored data is increasing every day. The researchers are continuously working on improvements over existing algorithms, that usually come along with the increasing number of datasets. A typical deep learning classifier consists of several layers of CNN between the input and output layers that allow for complex non-linear information processing [23]. Chua et al. [24] have made a comparison on three classical machine learning algorithms which are histogram of oriented gradient (HOG), exemplar-SVM and latent-SVM [25] to find their specific advantages, where they found out
that exemplar SVM is good for specificity measure. To overcome the issue of scale variance and complicated background, Yu et al. [15] introduced a detection approach using a fully convolutional neural network (FCNN) segmentation method and then detects the object through bounding box regressions, where the class is labelled by CNN classifier [26].

2.4. Network architecture
Krizhevsky et al. [16] proposed a deeper and wider CNN model compared to the original LeNet, where they won the most difficult ImageNet challenge (ILSVRC) in 2012. Their AlexNet managed to achieve state-of-art recognition accuracy against all the traditional machine learning algorithms. In 2013, Zeiler & Fergus [27] have made some improvements to AlexNet by tweaking the network parameters, where they won the 2013 ILSVRC. In 2014, Simonyan & Zisserman [28] have investigated the effect of convolutional on large scale images. GoogleNet, the winner of 2014 ILSVRC [29] introduced several parallel CNNs with different kernel sizes. The work in [30] has implemented residual networks that allow the architecture to be very deep with more than 100 layers. Zagoruyko & Komodakis [31] proposed a wider version of the residual network, while Xie et al. [32] introduced an aggregated residual transformation. Other than that, Szegedy et al. [33] introduced the concept of inception block with residual connection in Inception-v4 architecture. A feedforward CNN was developed by Huang et al. [34] that consists of densely connected CNN layers. The output of each layer is connected with all previous layers in one dense block.

3. Research Method

3.1. Dataset
In this paper, a dataset of ship images from satellite images is obtained from the Kaggle platform [35]. The satellite images provide views of the earth’s surface that includes agriculture, building, road and so on. These images are obtained from PlanetScope full visuals at the San Francisco Bay and San Pedro Bay areas of California. It consists of 4000 RGB images with the size of 80 x 80 pixels for two classes problem of “ship” and “non-ship”.

| Class       | Number of images |
|-------------|------------------|
| Ship        | 1,000            |
| Non-Ship    | 3,000            |

For the ship class, the images must contain the full appearance of a single ship. It comes with different sizes and orientations, where some atmospheric noises are also included. A non-ship class comprises one of these three components, which are 1) random samples of different land cover features such as building, water, vegetation, bare earth and so on 2) partial image of a ship and 3) noises that are caused by bright pixels or strong linear features.

| Class      | Images of ships |
|------------|-----------------|
| Ship       | ![Ship Images](image1.png) ![Ship Images](image2.png) ![Ship Images](image3.png) |
| No Ship    | ![No Ship Images](image4.png) ![No Ship Images](image5.png) ![No Ship Images](image6.png) |
3.2. DenseNet

DenseNet is deep learning architecture that connects each layer to every other layer in a feed-forward fashion. DenseNet main advantages are reduction in the vanishing-gradient problem and improvement in features propagation. For one dense block, each final layer obtains an additional input from the preceding layers, which is feed forwarded and combined through the concatenation process. DenseNet also found to be good for small numbers of training datasets as it reduces the over-fitting problem, especially if transfer learning is used.

![Figure 1. A 5-layer dense block with growth rate of k = 4.](image)

Instead of combining features through summation operator, DenseNet combines the feed forward features with the previous layer output through concatenation, where all features will be brought forward as it is. Thus, the nth layer will have n input that comprises of CNNs from the previous layers. Instead of L layer as in traditional architectures, DenseNet introduces \((L(L+1))/2\) connection for a \(L\)-layer network. Fig. 1 shows an example of the last layer concatenation process that takes input from the first four CNN layers.

4. Results and Discussion

4.1. Setup

In this paper, we design a classifier based on selected DenseNet architecture. This algorithm is compiled and programmed using Keras with Tensorflow. The algorithm is fine-tuned using several variations of layer's number and parameter as mentioned in Table III. Python platform is used to handle data and plotting the figures. Other modules that are also used in these experiments are numpy, open cv, pandas. The pre-trained DenseNet parameters are downloaded from the standard Keras library. Batch size will also be varied but it is limited to the capacity of our computer RAM and no graphics processing unit is used in the experiments.

| Table 3. Setting of the hyperparameters. |
|-----------------------------------------|
| Variables                  | Values                        |
|----------------------------|-------------------------------|
| Learning rates            | 0.001 and 0.0001              |
| Pooling layers            | Global Average               |
| Activation                | ReLu                          |
| Batch size                | 16 and 32                     |
| Optimizers                | Adam, Adamax, RMSprop and SGD |
4.2. Optimization

To train the DenseNet, an optimization algorithm is needed to update the parameters according to the loss function. It concerns on the differences between the model prediction and the label defined as ground truth. Adam optimizer is an algorithm for first-order gradient-based optimization of a stochastic objective function [36]. It is a computationally efficient optimizer and generally applicable to most types of data. Its memory usage is also relatively low. Finally, Adam also adds bias-correction and momentum to the basic RMSprop. While, Adamax is sometimes superior to Adam when model embedding is used. On the other hand, RMSprop focuses on the moving average of squared gradients. SGD optimizer is one of the earliest optimizers that does not implement momentum in calculating the update weights.

Table 4. Comparison of evaluation accuracy between different optimizers.

| Optimizer | Learning rate | Batch size | No. of epoch | Evaluation accuracy |
|-----------|---------------|------------|--------------|---------------------|
| Adam      | 0.0001        | 16         | 2            | 98.4375%            |
| Adamax    | 0.001         | 16         | 2            | 91.5000%            |
| RMSprop   | 0.001         | 16         | 2            | 84.0625%            |
| SGD       | 0.01          | 16         | 2            | 96.9380%            |

Figure 2. A performance comparison of ship classification between different optimizers.

According to Table 4, the best classification performance is obtained when Adam optimizer is used with 98.44% accuracy. This performance is obtained by using a learning rate of 0.0001, instead of the standard 0.001. The batch size is fixed to 16 for all types of optimizer. The least performed optimizer is RMSprop, which is relatively low in accuracy compared to the other optimizer.

4.3. Dataset division

Since a supervised learning approach is used in this paper, the training data must be labeled to train the algorithms. While testing data is the non-overlapping set of images used to measure the robustness of the classifier prediction. Thus, it is a set of data that has not been seen before. The dataset is split into three cases, which are either 7:3, 4:1 or 9:1 split ratio between the training and testing data. Results in Table 5 show that there are small performance differences between all the ratios, where the best accuracy is obtained by using 9:1 data split between the training and testing data. All other hyperparameters are fixed which includes a batch size of 16 and a learning rate of 0.0001.
Table 5. Comparison of evaluation accuracy between number of training and testing images (No. of epochs = 50).

| Training images | Testing images | Learning rate | Batch size | Evaluation accuracy |
|-----------------|----------------|---------------|------------|---------------------|
| 2,800           | 1,200          | 0.0001        | 16         | 99.0833%            |
| 3,200           | 800            | 0.0001        | 16         | 99.5000%            |
| 3,600           | 400            | 0.0001        | 16         | 99.7500%            |

Figure 3. A Performance comparison of ship classification by training and testing images.

4.4. Batch size
Batch size is a hyperparameter that defines the number of samples processed for one training iteration. For example, if we have a batch size of 32 samples, the algorithm will process the first 32 images (from 1st to 32nd) for the first iteration. Next, the algorithm will fetch the second 32 images (from 33rd to 64th) for the second iteration and it will continue until it processed all the images for an epoch operation. Table 6 shows the performance comparison between batch sizes of 16 and 32. A batch size of 16 delivers a better accuracy of 88.73% compared to a batch size of 32 with just 80.65% accuracy. Hence, a bigger batch size does not always produce a better classification accuracy.

Table 6. Comparison of accuracy evaluation between different batch sizes.

| Training images | Testing images | Learning rate | Batch size | No. of epoch | Evaluation accuracy |
|-----------------|----------------|---------------|------------|--------------|---------------------|
| 2,400           | 1,600          | 0.0001        | 32         | 2            | 80.6538%            |
| 2,800           | 1,200          | 0.0001        | 16         | 2            | 88.7308%            |

4.5. Learning rates
Learning rate effects on the weights of the gradient error that got updated. Specifically, it controls the amount of errors that the weights of the model will be included for each time it is updated. Two learning rates are explored, which are 0.001 and 0.0001. The smaller learning rate of 0.0001 produces a much better accuracy compared to the 0.001 learning rate. A bigger learning rate might cause local optima issue but a smaller learning rate will make the training process longer.
## 5. Conclusion

In this paper, we have proposed a ship detection and classification method using DenseNet for remote sensing images. DenseNet is an extremely sophisticated algorithm that is able to classify the ships with more than 90% accuracy. The best combination of hyperparameter is obtained when a batch size of 16 and a learning rate of 0.0001 are used. Adam optimizer is also found to be the best optimization method for our ship detection using DenseNet. For better results in the future, the number of images in the dataset can be increased to cover a more variety of ship types.

## Acknowledgments

The authors would like to acknowledge funding from Universiti Kebangsaan Malaysia (GUP-2019-008) and Ministry of Education Malaysia (FRGS/1/2019/ICT02/UKM/02/1). Graphic Processing Unit used in training the deep learning model is donated by NVIDIA Corp. (KK-2019-005).

## References

[1] Crisp D J and Keevers T 2010 Comparison of ship detectors for polarimetric SAR imagery *IEEE OCEANS*10 pp 1-8
[2] Vesecky J F, Laws K E and Paduan J D 2009 Using HF surface wave radar and the ship Automatic Identification System (AIS) to monitor coastal vessels *IEEE International Geoscience and Remote Sensing Symposium* pp 761-4
[3] Zulkifley M A, Abdani S R and Zulkifley N H 2019 Pterygium-Net: a deep learning approach to pterygium detection and localization *Multimedia Tools and Applications* 78 34563-84
[4] Zulkifley M A and Trigoni N 2018 Multiple-model fully convolutional neural networks for single object tracking on thermal infrared video *IEEE Access* 6 42790-9
[5] Abdani S R and Zulkifley M A 2019 DenseNet with spatial pyramid pooling for industrial oil palm plantation detection *Int. Conf. Mechatronics, Robot. Syst. Eng.* pp 134–8
[6] Abdani S R, Zulkifley M A and Hussain A 2019 Compact convolutional neural networks for pterygium classification using transfer learning *IEEE International Conference on Signal and Image Processing Applications* pp 140-3
[7] Zulkifley M A 2019 Two streams multiple-model object tracker for thermal infrared video *IEEE Access* 7 32383-92
[8] Zulkifley M A, Mohamed N A and Zulkifley N H 2019 Squat angle assessment through tracking body movements *IEEE Access* 7 48635-44
[9] Rostami M, Kolouri S, Eaton E and Kim K 2019 Deep transfer learning for few-shot SAR image classification *Remote Sensing* 11 1-19
[10] Kanjir U, Greidanus H and Ostir K 2018 Vessel detection and classification from spaceborne optical images: A literature survey *Remote Sensing of Environment* 207 1-26
[11] Lin Y, He H, Tai H, Chen F and Yin Z 2017 Rotation and scale invariant target detection in optical remote sensing images based on pose-consistency voting *Multimed. Tools Appl.* 76 14461–83
[12] Huang S, Xu H and Xia X 2015 A remote sensing ship recognition using random forest *Proceedings of Science* pp 1-11
[13] Yokoya N and Iwasaki A 2015 Object detection based on sparse representation and Hough voting for optical remote sensing imagery *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8 2053-62
[14] Prasad D K, Prasath C K, Rajan D, Rachmawati L, Rajabaly E and Quek C 2016 Challenges in

### Table 7. Comparison of evaluation accuracy between different learning rates.

| Training images | Testing images | Learning rate | Batch size | No of epoch | Evaluation accuracy |
|-----------------|---------------|---------------|------------|-------------|---------------------|
| 2,400           | 1,600         | 0.001         | 16         | 2           | 89.6875%            |
| 2,400           | 1,600         | 0.0001        | 16         | 2           | 98.4375%            |
video based object detection in maritime scenario using computer vision CoRR pp 1-6

[15] Yu J G, Xia G, Deng J and Tian J 2015 Small object detection in forward-looking infrared images with sea clutter using context-driven Bayesian saliency model Infrared Phys. Technol. 73 175-83

[16] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks Advances in Neural Information Processing Systems pp 1-9

[17] Rawat W and Wang Z 2017 Deep convolutional neural networks for image classification: A comprehensive review Neural Computation 29 2352-449

[18] Tang J, Deng C, Huang G and Zhao B 2015 Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine IEEE Trans. Geosci. Remote Sens. 53 1174-85

[19] Zhang B, Su J, Xiong D, Lu Y, Duan H and Yao J 2015 Shallow convolutional neural network for implicit discourse relation recognition Conference on Empirical Methods in Natural Language Processing pp 2230–35

[20] Borji A, Cheng M M, Hou Q, Jiang H and Li J 2019 Salient object detection: A survey Computational Visual Media 5 117–50

[21] Liu Y, Cui H, Kuang Z and Li G 2017 Ship detection and classification on optical remote sensing images using deep learning ITM Web Conf. 12 pp 1-6

[22] Turner J T, Gupta K M and Aha D 2017 SPARCNN: Spatially related convolutional neural networks IEEE Applied Imagery Pattern Recognition Workshop pp 1-6

[23] Lecun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44

[24] Chua M, Aha D W, Auslander B, Gupta K and Morris B 2013 Comparison of object detection algorithms on maritime vessels Tech. Rep. AIC–14-041, Nav. Res. Lab. Navy Cent. Appl. Res. Artif. Intell. pp 1-11

[25] Mohamed N A, Zulkiifley M A, Zaki W M D W and Hussain A 2019 An automated glaucoma screening system using cup-to-disc ratio via Simple Linear Iterative Clustering superpixel approach Biomedical Signal Processing and Control 53 p 101454

[26] Audebert N, Saux B and Efèvre S 2017 Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images Remote Sens. 9 pp 1-18

[27] Zeiler M D and Fergus R 2014 Visualizing and understanding convolutional networks Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) pp 818-33

[28] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition International Conference on Learning Representations pp 1-14

[29] Szegedy C et al. 2014 Going deeper with convolutions IEEE Conference on Computer Vision and Pattern Recognition pp 1-9

[30] He K, Zhang X, Ren S, and Sun J 2015 Deep residual learning for image recognition IEEE Conference on Computer Vision and Pattern Recognition pp 770-778

[31] Zagoruyko S and Komodakis N 2016 Wide residual networks British Machine Vision Conference pp 1-12

[32] Xie S, Girshick R, Dollar P, Tu Z and He K 2017 Aggregated residual transformations for deep neural networks IEEE Conference on Computer Vision and Pattern Recognition pp 5987-95

[33] Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z 2016 Rethinking the Inception architecture for computer vision IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp 2818-26

[34] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks IEEE Conference on Computer Vision and Pattern Recognition pp 2261-69

[35] https://www.kaggle.com/rhammell/ships-in-satellite-imagery

[36] Kingma D P and Ba J L 2015 Adam: A method for stochastic optimization 3rd International Conference on Learning Representations pp 1-15