

Ship Identification on Satellite Image Using Convolutional Neural Network and Random Forest

Endang Anggiratih¹, Agfianto Eko Putra*²

¹Master Program of Computer Science; FMIPA UGM, Yogyakarta, Indonesia

²Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: ¹endang.anggiratih@mail.ugm.ac.id, *²agfi@ugm.ac.id

Abstrak

Identifikasi kapal pada citra satelit dapat dimanfaatkan untuk pengelolaan perikanan, pemantauan kegiatan penyelundupan, layanan lalu lintas kapal, maupun perang angkatan laut. Namun citra satelit resolusi tinggi juga membuat segmentasi kapal dengan latar-belakang menjadi sulit, sehingga untuk menangani hal tersebut dibutuhkan fitur-fitur handal sehingga dapat diidentifikasi dengan cukup baik antara kapal besar, kapal kecil dan bukan kapal. Metode Convolutional Neural Network (CNN), yang memiliki keunggulan dapat mengekstraksi fitur secara otomatis serta menghasilkan fitur-fitur handal yang memudahkan identifikasi kapal. Penelitian ini mengkombinasikan arsitektur CNN ZFNet dengan metode Random Forest. Pelatihan dilakukan dengan tujuan mengetahui akurasi dari lapisan-lapisan ZFNet agar menghasilkan fitur yang terbaik, yang ditandai dengan akurasi yang tinggi, dipadukan dengan metode Random Forest. Pengujian kombinasi metode ini dilakukan dengan dua parameter yaitu ukuran batch dan jumlah pohon. Hasil pengujian identifikasi kapal besar dengan akurasi 87.5%, dan kapal kecil dengan akurasi tidak sampai 50%.

Kata kunci—ekstraksi fitur, identifikasi kapal, CNN, ZFNet, Random Forest

Abstract

Ship identification on satellite imagery can be used for fisheries management, monitoring of smuggling activities, ship traffic services, and naval warfare. However, high-resolution satellite imagery also makes the segmentation of the ship difficult in the background, so that to handle it requires reliable features so that it can be identified adequately between large vessels, small vessels and not ships. The Convolutional Neural Network (CNN) method, which has the advantage of being able to extract features automatically and produce reliable features that facilitate ship identification. This study combines CNN ZFNet architecture with the Random Forest method. The training was conducted with the aim of knowing the accuracy of the ZFNet layers to produce the best features, which are characterized by high accuracy, combined with the Random Forest method. Testing the combination of this method is done with two parameters, namely batch size and a number of trees. The test results identify large vessels with an accuracy of 87.5% and small vessels with an accuracy of not up to 50%.

Keywords—features extraction, ships identification, CNN, ZFNet, Random Forest

1. INTRODUCTION

Satellite imagery makes it easy to recognize certain objects on the surface of the earth, for example buildings, roads, plantations, rice fields, pedestrians, and classification of ships at sea. Archipelagic countries surrounded by the oceans make detection and classification of ships as very important things to consider. Detection and classification of ships can be used for fisheries management, supervision of smuggling activities, ship traffic services, and sea wars. [1,2].

High resolution satellite images produce images that have more detailed information [3]. However, high resolution images make the background part difficult to separate so that it will increase processing time and even cause many false alarms. To deal with the complexity of high-resolution images, the most important requirements are reliable features, which are able to distinguish objects from non-objects, while the other main requirement is the accuracy of the method used [4].

The research carried out by [5,6] applied the Threshold method as the background segmentation with ships and was able to separate the background with the ratio indicated by the threshold and the blue ribbon. The sea area usually has a stationary gray distribution with low and gray scale variations, different from artificial objects that are shown through histograms with threshold segmentation.

Ship and non-ship segmentation using the threshold alone is not enough because it will experience difficulties in separating existing vessels at the port, because the color and shape of lines in ports and vessels have similarities, therefore it needs to be combined with machine learning to improve efficiency and reliability, especially deep learning [7] uses the convolutional neural network (CNN) method which is able to automatically extract features properly. But CNN, like other deep learning methods, has weaknesses in the training process that take a long time, especially when using multiple layers.

Research using other machine learning methods was carried out by [8] who applied the Random Forest method for classification, which had a fairly high accuracy compared to the support vector machine method (SVM), but Random Forest methods took a long time to predict if a large number of trees were needed. The disadvantages of several methods of deep learning and machine learning that produce high accuracy usually require long training time. Therefore the use of a combination of deep learning methods and machine learning can be applied to overcome long periods of time during training and are expected to produce high accuracy.

2. METHODS

2.1 Research Flow

The steps taken in this study include the stage of shooting, preprocessing, preparation of datasets, ZFNet-Random Forest training, ZFNet-Random Forest testing, and analysis of research results. Figure 1 shows a chart of the research process. The initial stage of the research is data collection, then pre-processing to produce a dataset by detecting prospective vessels. The dataset consists of three images, namely images of large ships, small vessels, and non-ships. Datasets are designed separately for each class with different sizes. After the dataset is ready, the next process is the training and testing process. The final stage is an analysis of research results to draw conclusions.

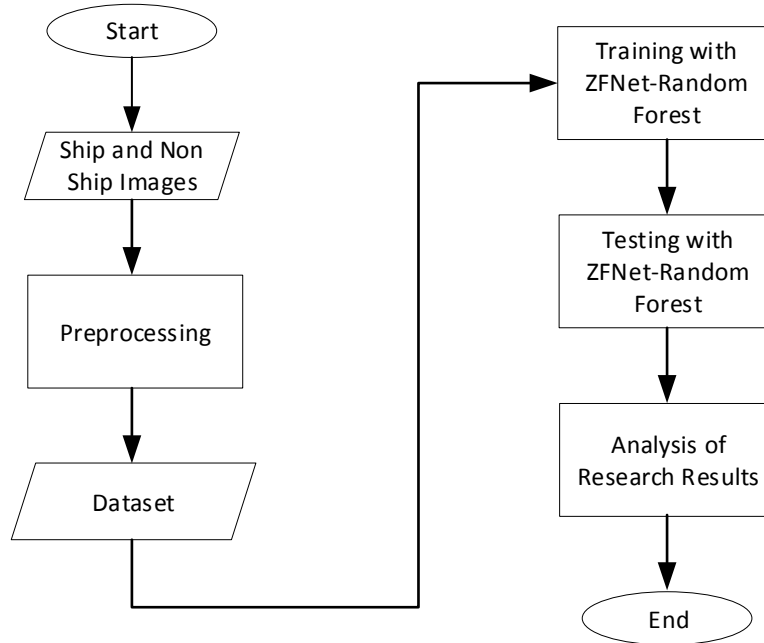


Figure 1 Research Flowchart

2. 1.1 Images Pre-processing

The initial stage is image pre-processing, at this stage the RGB ship image will be detected to be introduced as a candidate ship using the HOG-SVM algorithm, using LUV color space parameters. The results of identification of prospective ship images are then divided into three classes, pictures of large ships, pictures of small ships, and non-ship pictures. Previously, this algorithm was trained by inputting an 80x80 as many as 2,800 images divided into two classes, namely the ship class and the non-ship class.

2. 1.2 CNN Method

Convolutional networks, known as convolutional neural networks (CNN), are special types of neural networks for processing data that have mesh or grid-like topologies. Convolutional neural network names indicate that the network uses convolution mathematical operations, which are linear operations. Thus Convolutional Network is a neural network that uses minimal convolution in one layer [9]. Technically, convolutional networks are architectures that can be trained and consist of several stages. Inputs and outputs from each stage are some arrays called feature maps. Example of a gray scale image, the input is a two-dimensional matrix. The output of each stage consists of three layers, namely convolution, activation, and unification layers.

2. 1.3 Convolutional Layer

The convolutional layer carries out convolution operations on the output from the previous layer. This layer is the main process that underlies CNN, which is to apply functions to other functions over and over again. Convolution operations are imposed on the function $x(t)$ with weights (or often called kernels) $w(t)$, written with operators $*$, or written as $x*w$, as shown in Equation 1.

$$s(t) = (x * w) = \sum_{a=-\infty}^{-\infty} x(a) * w(t - a) \quad (1)$$

where $s(t)$ is a function of convolution operations, t is a time variable, and a is a constant. In

digital image processing, convolution is understood by moving an $m \times n$ -sized K kernel to an $i \times j$ -sized image, then taking the number of copies of the image and kernel values. The term convolution is almost the same as the term correlation. In machine learning applications, the two terms are considered the same, so when the convolution is done, the kernel can be reversed first or unnecessary. Formally, convolution in sxt , an image size $I(sxt)$, with a kernel of size $m \times n$, $K(m \times n)$, can be expressed through equations 2 and 3.

$$s[i, j] = (I * K)[i, j] = \sum_m \sum_n I[m, n]K[i - m, j - n] \quad (2)$$

$$s[i, j] = (I * K)[i, j] = \sum_m \sum_n I[m, n]K[i - m, j - n] \quad (3)$$

2. 1.4 Pooling Layer

Pooling layer is the process of reducing the size of image data. In processing, integration also aims to increase the invariance of feature positions. In most CNN, the pooling method also called the subsampling method used is max pooling. Max pooling divides the output from the convolution layer into a number of small grids, which then take the maximum value from each grid to arrange the matrix of the reduced image, as shown in Figure 2. Figure 2 shows grids in red, green, yellow, and blue (left side) is a box group that will select the maximum value. So the results of the process can be seen on a smaller grid set (right). This process ensures that the features obtained will be the same even though the object image is translated.

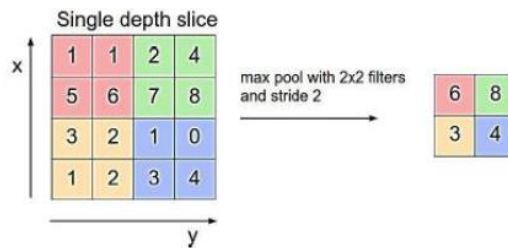


Figure 2 Max pooling

2. 1.5 Fully Connected Layer

The neurons are fully connected to all activations, which are connected to the previous layer, this layer is always placed behind the layer, so there is no convolutional layer after the layer is fully connected. Used for the classification process using matrix multiplication and offset bias.

2. 1.6 ZFNet Architecture

The researchers competed to develop CNN architecture with the aim of getting good performance for complex models. Compared to the previous CNN architecture, such as LeNet, many researchers concentrated on progress in performance. In particular, Zeiler and Fergus (2014) made a detailed analysis of optimality and the means to correct it based on the statement: "There is no clear understanding of why CNN works so well, or how CNN can be improved. There is still little insight into internal operations and behavior this complex model, or how CNN achieved such good performance. From a scientific point of view, this is very unsatisfactory "[10]. The architecture created by Zeiler and Fergus was named ZFNet, where the architecture achieved a big error rate of 14.8% compared to the previous architecture. The ZFNet architecture is shown in Figure 3.

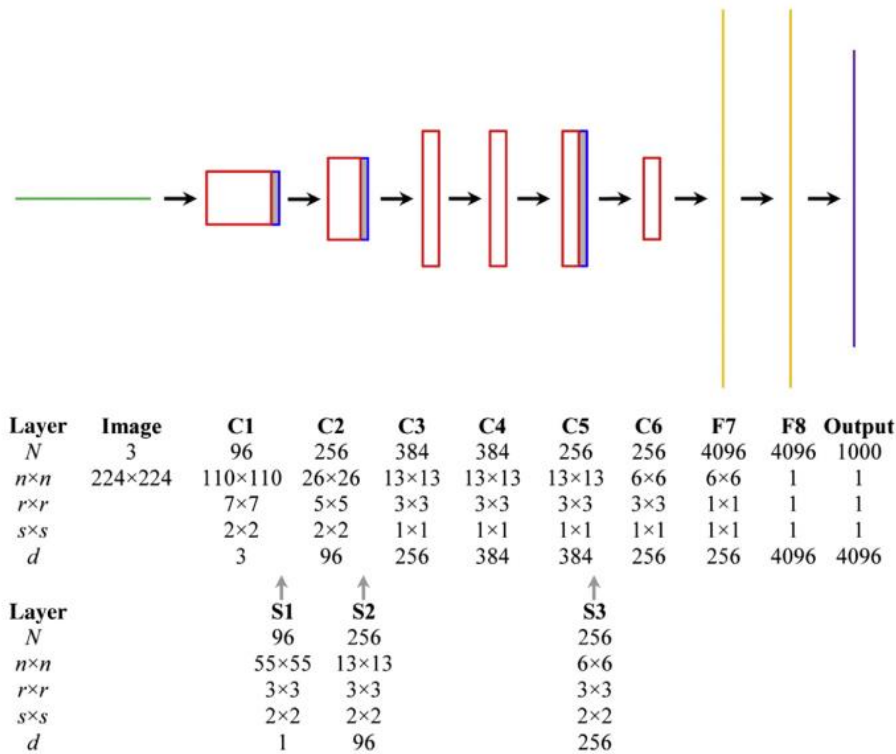


Figure 3 ZFNet Architecture

This study uses a combination of two methods, namely CNN (ZFNet) and Random Forest, as shown in Fig. 4. The first part is feature extraction with ZFNet which has 6 convolutional layers called C where the following numbers show the order of layers, 3 layers union is called S and 2 layers are fully connected called F. While the second part is the identification stage of the ship using Random Forest. The ZFNet layer has a size that is not the same between several layers as shown in Table 1.

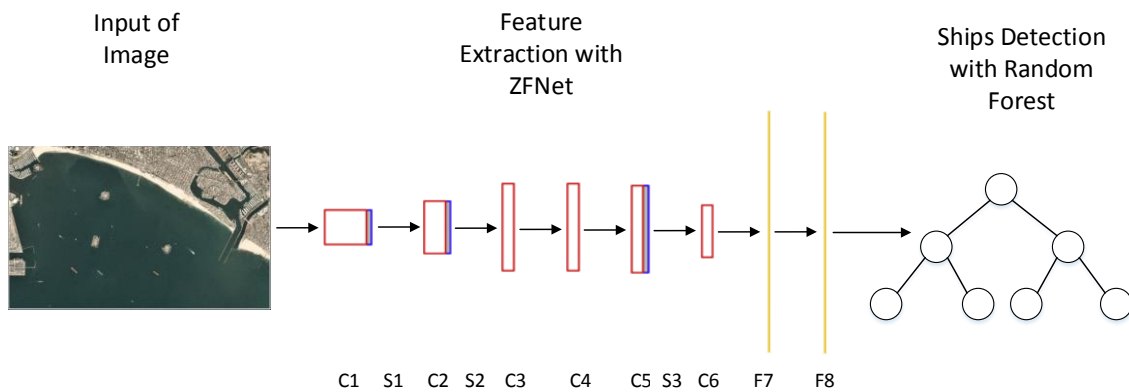


Figure 4 ZFNet-RandomForest Architecture

2. 1.7 Random Forest Method

Random Forest is the development of the CART method, namely by setting the bootstrap method and random feature selection. Random forest is a classification method that contains a number of decision trees, first proposed by Breiman in 2001. Random forests can be used for various types of response variables such as continuous, discrete, survival data and multivariate combination data [11]. In addition, there are no assumptions that must be fulfilled in random forests. This method can estimate various forms of functions that are formed between

response and explanatory variables and make it easier to determine complex nonlinear relationships that might be difficult to find without certain specifications and without using certain standard methods. In essence, random forests can and are able to detect various interactions between responses and predictors. The flexibility of random forests makes this method very useful as a method of data exploration. Ordinary random forests are also referred to as ensemble methods or combined methods. It is called a combined method because it is formed from a small model, but the prediction results are determined by combining all outputs on a small model or what are called sub-models [12].

Table 1 ZFNet Layer Size

Layer Name	Size			Note
	Layer	Filter	Stride	
C1	110x110	7x7	2x2	<i>Convolution 1st</i>
S1	55x55	3x3	2x2	<i>Subsampling 1st (Pooling)</i>
C2	26x26	5x5	2x2	<i>Convolution 2nd</i>
S2	13x13	3x3	2x2	<i>Subsampling 2nd (Pooling)</i>
C3	13x13	3x3	1x1	<i>Convolution 3rd</i>
C4	13x13	3x3	1x1	<i>Convolution 4th</i>
C5	13x13	3x3	1x1	<i>Convolution 5th</i>
S3	6x6	3x3	2x2	<i>Subsampling 3rd (Pooling)</i>
C6	6x6	3x3	1x1	<i>Convolution 6th</i>
F7	6x6	1x1	1x1	<i>Fully Connected 1st</i>
F8	1	1	1	<i>Fully Connected 2nd</i>

2. 2 Feature Extraction

The feature extraction stage is done using CNN specifically with the ZFnet architecture. Input images consist of three classes: large ships, small vessels, and non-vessels. The input image varies in size due to the results of candidate identification, therefore it needs to be normalized first to be 80×80 and the input image is converted to gray scale, to reduce the time calculation in the feature extraction process.

2. 3 Division of the Dataset

The dataset is divided into 5 scenarios to get the highest accuracy value. The total image data is 420 where there are 3 classes, 77 large ship classes, 37 small ship classes, and 316 non-class images. The distribution of the dataset is used to obtain overall accuracy, because there are not too many datasets used, the dataset is divided into 20%, 40%, 60%, 80%, and 100% of the drawing dataset for training.

2. 4 Ship Images Identification

Before carrying out the classification stage that needs to be done is the CNN training stage using the ZFnet architecture combined with the Random Forest method. The training used 420 image data as a result of candidate ship detection, while training for candidate ship detection used 2,800 image data consisting of two classes namely ship and non-ship class, ship class including large vessels and small vessels. CNN training will produce the best models with high training accuracy, so as to provide the best feature extraction results. Fig. 5 shows the stages of the identification process.

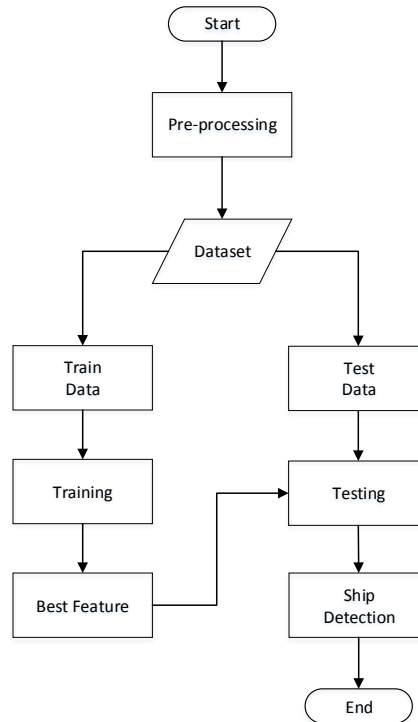


Figure 5 Stages of the Identification Process

3. RESULTS AND DISCUSSION

3.1 Testing Feature Extraction

The first test was conducted to determine the accuracy of the training in the three layers to be tested, namely the convolution layer 6, the fully connected layer 7, and the fully connected layer 8. The layer tested only the last three layers on ZFNet due to maintaining the ZFNet architecture itself. Testing is done using a number of different datasets. Training datasets are 20%, 40%, 60%, 80% and 100%, testing is done 5 times according to the variant number of datasets. The test results are shown in Table 1.

Table 1 Feature extraction results on three ZFNet layers

Used of Training Data	Layer Accuracy		
	Convo 6	FC 7	FC 8
20%	99,52%	99,54%	99,49%
40%	99,48%	99,51%	99,49%
60%	99,50%	99,52%	99,53%
80%	99,52%	99,52%	99,53%
100%	99,52%	99,52%	99,2%

Table 1 shows the difference in accuracy that is not too significant. The highest accuracy of each layer has been tested compared to other methods. In the fully connected layer 7 the highest accuracy value in the dataset is 20% and 40% with an accuracy value of 99.54% and 99.51%. The highest accuracy value in the fully connected 8 layer is in the dataset 60% and 80% with an accuracy value of 99.53% and 99.53%. Accuracy value is the same as high in convolution layer 6, fully connected 7, and fully connected 8 when dataset is 100%.

3.2 Testing ZFNet-Random Forest Performance

Tests carried out at the ship image classification stage using the ZFNet-Random Forest method. The testing phase is based on the highest value of training accuracy in fully connected 7 and fully connected 8 layers compared to the ZFNet-SVM method, Table 2 shows the results of ZFnet-Random Forest Convolution layer 6 training performance.

Table 2. ZFnet-Random Forest Convolution layer 6 training performance

ZFNet Layer	Number of Random Forest tree (n_estimator)	Batch size ZFNet	Training time		Accuracy
			ZFNet feature extraction	Random Forest identification	
Convo6	100	100	0.55	480.5	98.9%
Convo6	100	200	0.67	473.0	98.9%
Convo6	100	300	0.79	497.3	99.3%
Convo6	100	400	1.03	496.1	99.5%
Convo6	200	100	0.95	496.7	98.9%
Convo6	200	200	1.19	4.66.1	98.9%
Convo6	200	300	1.41	494.8	99.3%
Convo6	200	400	1.87	498.9	99.5%
Convo6	300	100	1.36	480.8	98.9%
Convo6	300	200	1.72	482.2	99.0%
Convo6	300	300	2.14	487.5	99.3%
Convo6	300	400	2.59	488.5	99.5%

Table 2 shows that the more the number of trees and the size of the batch the longer the training time is, but the accuracy is higher. The same results were obtained for 8 fully connected layers in Table 4, but good results were obtained in 7 fully connected layers in Table 3 where the batch size was 400 chips and the number of trees was from 100 to 300, with the result being accuracy of 99.0 %.

Table 3 ZFnet-Random Forest fully connected layer 7 training performance

ZFNet Layer	Number of Random Forest tree (n_estimator)	Batch size ZFNet	Training time		Accuracy
			ZFNet feature extraction	Random Forest identification	
FC7	100	100	0.56	492.5	98.9%
FC7	100	200	0.67	502.9	99.0%
FC7	100	300	0.79	494.1	99.3%
FC7	100	400	1.01	489.9	99.5%
FC7	200	100	0.95	490.8	98.9%
FC7	200	200	1.18	490.2	99.0%
FC7	200	300	1.42	506.7	99.3%
FC7	200	400	1.86	487.0	99.5%
FC7	300	100	1.26	492.9	99.0%
FC7	300	200	1.68	502.1	99.1%
FC7	300	300	2.03	490.4	99.3%
FC7	300	400	2.61	491.8	99.5%

3.3 Testing ZFNet-Random Forest Identification

The identification stage uses 8 satellite images from the San Francisco port, carried out for the three deepest layers. The results are shown in Table 5 for identification of large vessels and Table 6 for identification of small vessels.

Table 4 ZFnet-Random Forest fully connected layer 8 training performance

ZFNet Layer	Number of Random Forest tree (n_estimator)	Batch size ZFNet	Training time		Accuracy
			ZFNet feature extraction	Random Forest identification	
FC8	100	100	0.54	560.8	98.8%
FC8	100	200	0.67	564.5	98.9%
FC8	100	300	0.78	577.3	99.3%
FC8	100	400	1.04	578.5	99.5%
FC8	200	100	0.96	553.5	98.9%
FC8	200	200	1.18	547.7	98.9%
FC8	200	300	1.53	562.1	99.3%
FC8	200	400	1.78	571.2	99.5%
FC8	300	100	1.35	542.3	98.9%
FC8	300	200	1.70	539.4	99.0%
FC8	300	300	2.14	544.2	99.3%
FC8	300	400	2.59	553.4	99.5%

Tabel 5 Big vessel identification testing

Name	Original	20%-FC7	40%-FC7	60%-FC8	80%-FC8
lb_1.png	5	5	5	5	5
lb_2.png	11	11	11	11	11
lb_3.png	7	9	7	10	6
lb_4.png	3	3	3	3	3
sfbay_1.png	9	9	9	9	9
sfbay_2.png	8	8	8	8	8
sfbay_3.png	9	9	9	9	9
sfbay_4.png	1	1	1	1	1

Tabel 6 Small vessel identification testing

Nama Citra	Data Asli	20%-FC7	40%-FC7	60%-FC8	80%-FC8
lb_1.png	5	5	5	5	5
lb_2.png	10	9	9	9	9
lb_3.png	8	5	6	4	6
lb_4.png	11	9	9	9	9
sfbay_1.png	5	4	4	4	4
sfbay_2.png	1	1	1	1	1
sfbay_3.png	1	1	1	1	1
sfbay_4.png	9	9	8	10	10

The average accuracy for identification of large vessels is 87.5% while for small vessels low accuracy is not up to 50.0% because the image of the small ship is very small, so that in the test it is considered non-ship.

4. CONCLUSIONS

The conclusions from the results of this study are as follows: accuracy produced in convolution 6, fully connected 7, and fully connected 8 layers did not produce a significant difference in accuracy. The use of the Random ZFNet-Forest method combination uses two parameters, namely batch size and number of trees (n_estimator) resulting in 99.0% accuracy obtained from the FC7 layer. The classification stage of large vessels is able to recognize with an accuracy of 87.5%, while for small vessels it is very low, because the ship's image is so small that it is difficult to distinguish small vessels from non-ships.

ACKNOWLEDGMENTS

This research is supported by ELINS laboratory in the department of computer sciences and electronics of Universitas Gadjah Mada.

REFERENCES

- [1] W. Guo, X. Xia, and W. Xiaofei, "A remote sensing ship recognition method based on dynamic probability generative model", *Expert System with Applications*, vol. 41, issue 14, Oct 2014 [online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417414001705>. [Accessed: 22-Mar-2017]
- [2] L. Chunyan, Z. Huanxin, S. Hao, and Z. Shilin, "Combing Rough Set and RBF Neural Network for Large-Scale Ship Recognition in Optical Satellite Images", *IOP Conference Series: Earth and Environmental*, vol 17, con 1, 2014 [Online]. Available: <http://iopscience.iop.org/article/10.1088/1755-1315/17/1/012218/pdf>. [Accessed: 22-Mar-2017]
- [3] G. Liu, Y. Zhang, X. Zheng, X. Sun, K. Fu, and H. Wang, "A new method on inshore ship detection in high-resolution satellite images using shape and context information" *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, March 2014
- [4] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship Detection in High-Resolution Optical Imagery Based on Anomaly Detector and Local Shape Feature", *IEEE Transaction on Geoscience and Remote Sensing*, vol. 52, issue 8, Aug 2014 [online]. Available: <http://levir.buaa.edu.cn/publications/TGRS-shipDetectionRXHOG.pdf>. [Accessed: 22-Mar-2017].
- [5] X. Leng, K. Ji, K. Yang, and H. Zou, "A Bilateral CFAR Algorithm for Ship Detection in SAR Images" *IEEE Geoscience and Remote Sensing*, vol. 12, issue 7, July. 2015.
- [6] J. Xu, X. Sun, D. Zhang, and K. Fu, "Automatic Detection of Inshore Ships in High-Resolution Remote Sensing Images Using Robust Invariant Generalized Hough Transform", *IEEE Transaction on Geoscience and remote sensing*, vol. 11, no. 12, Dec. 2014.
- [7] Y. Liu, M. Zhang, and Z. Guo, "SAR Ship Detection Using Sea-Land Segmentation-Based Convolutional Neural Network", *IEEE International Workshop on Remote Sensing with Intelligent Processing*, June 2017.
- [8] J. Frost, T. Geisler, and A. Mahajan, "Monitoring Illegal Fishing through Image Classification", Stanford.edu [Online]. Available: <http://cs229.stanford.edu/proj2016/report/FrostGeislerMahajan-MonitoringIllegalFishingThroughImageClassification-report.pdf>. [Accessed: 15-june-2017].
- [9] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning", MIT Press, 2016 [Online]. Available: www.deeplearningbook.org [Accessed: 12-Sep-2016].
- [10] E.R. Davies, "Deep Learning", *Computer Vision* 5th Edition, page: 453-493, Elsevier.
- [11] S.W. Eric, J.K. Betty, G.C. Mark, and N. Diane, "A Random Forest Approach to Predict the Spatial Distribution of Sediment Pollution in an Estuarine System", *journals.plos.org*, 2017 [Online]. Available: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179473> [Accessed: 12-Jul-2018].
- [12] W. Chen, X. Xie, J. Peng, H. Shahabi, H. Hong, D.T. Bui, Z. Duan, S. Li, and A. Zhu, "GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method", *CATENA*, vol. 164, May 2018, pages: 135-149 [Online] Available: <https://www.sciencedirect.com/science/article/pii/S0341816218300122> [Accessed: 12-Jul-2018]